Perception of Gaze and Head Direction in Groups of Faces

Joseph Florey

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Queen Mary
University of London

Statement of Originality

I, Joseph Florey, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of collaboration and publications:

Content in chapter 2 has previously been published in its entirety in Journal of Experimental

Psychology: Human Perception and Performance. Reference:

Florey, J., Dakin, S. C., Clifford, C. W. G., & Mareschal, I. (2015). Peripheral processing of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1084–1094. https://doi.org/10.1037/xhp0000068

Content in Chapter 3 has previously been published in Scientific Reports. Reference:

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*. https://doi.org/10.1038/srep32210

Abstract

Gaze direction and head rotation are powerful cues that inform humans about another person's attention, intentions and even emotion. Previous research has focused on understanding how people make judgements about individual faces in direct view. However in everyday life, people are often presented with groups of faces and need to judge where the attention of that group is directed, such as in group conversations or when giving presentations. This thesis presents research whose aim is to better understand how gaze direction and head rotation are perceived in the visual periphery and in groups.

First, observers' perception of gaze deviation in the visual periphery was tested, using psychophysical methods and modelling. The results showed that observers' ability to judge gaze perception is severely limited, and that observers' judgements are severely biased by head rotation in the visual periphery. Second, observers' ability to perceive the average gaze or head direction of a group of spatially distributed faces was investigated. This was done using equivalent noise analysis, a technique which gives estimates for observers' *internal noise* (how certain they are in their judgements of any individual face) and their *effective sample size* (how many faces they are able to combine into their average). The findings revealed that head rotation was averaged with less uncertainty and greater effective sample size than gaze deviation, suggesting that observers can more precisely and efficiently pool information about head rotation than gaze. Finally, averaging of heads and gaze stimuli presented in temporal sequences was analysed using the same equivalent noise technique and compared to spatial averaging. In sequences, the differences in processing between head and gaze direction disappear, suggesting that poor peripheral perception of gaze is the limit on our averaging of gaze cues.

Contents

Chapter 1 Introduction

1. Why study gaze perception?

In order to function in any social setting, people need to be able to understand and interpret non-verbal social cues from others. These cues can take the form of gestures, facial expressions or body language, however one of the most important cues comes from knowing where another persons' attention is directed (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Driver et al., 1999; Farroni, Csibra, Simion, & Johnson, 2002), which is achieved primarily by combining information from a person's gaze deviation and head rotation. The ability to rapidly and accurately judge where someone is looking is crucial for social interaction and can influence social processing; from interpreting emotional expressions to inducing feelings of anxiety. This section will discuss the evidence that demonstrates the importance of gaze perception and highlight the need to better understand gaze processing under all viewing conditions.

1.1. Importance of the eye region

Yarbus (1967) provided some of the first research showing the importance of the eye region in human interactions. In one of the earliest eye tracking studies, Yarbus presented complex scenes which contained both people and objects, and tracked the direction of observers' gaze as they looked at the picture (Fig 1). In a condition where observers were instructed to simply observe the image with no prior instruction, the areas that were most often fixated were the eye regions of all the individuals in the room (fig 1b). This effect was even stronger when the participants were asked to judge a social aspect of the picture, such as whether each person was familiar with a guest who had just entered the room (fig 1c). In a similar, more recent study, Henderson, Williams and Falk (2005) presented participants with faces and asked them to remember their identities. In one condition, participants' held fixation and could not make eye movements across the image, whereas in the other they were free to look where they chose. In the free viewing condition, participants fixated the eyes with much greater frequency than any other feature on the face; furthermore, participants' memory for the identity of the faces was significantly better in the free view condition. These results show that even without any prior cue, people will look to others' eyes above any other location and that this can enhance face perception and memory.

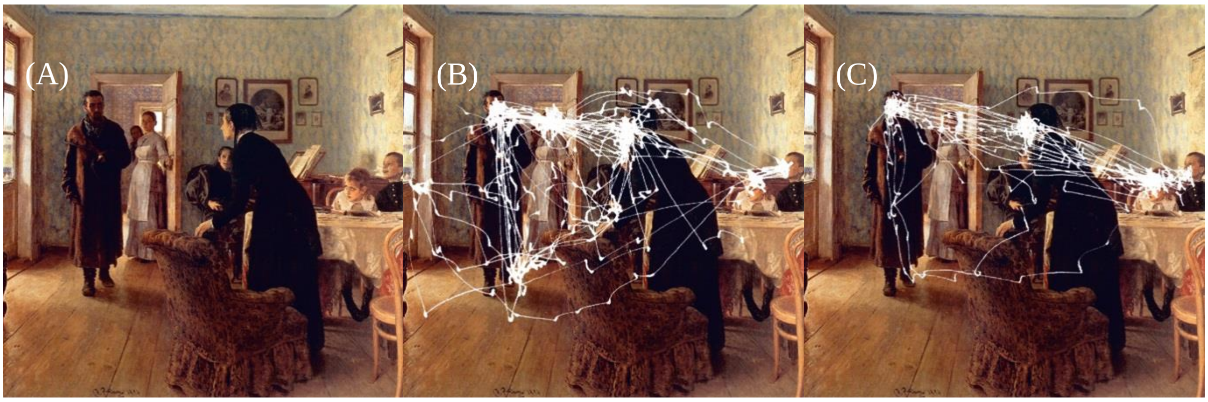

Figure 1:1 – Adapted from the stimuli in Yarbus's (1976) showing combined participants scan paths when viewing a complex scene.

(A) The basic image. (B) Scan paths when freely fixating the image. (C) Scan path when participants were asked to judge if those in the room knew the person who has entered.

There is a suggestion that human's eyes have evolved in such a way as to be highly salient and easily discriminable. Emery (2000) observed that, compared to most monkey species, humans have flatter faces, higher cheek bones and eyebrows which frame the eye region, suggesting this region is more important in humans than any other animal. Similarly, the human eye contains much greater contrast between the iris and sclera (white) of the eye than all other mammals, whose irises tend to make up the majority of the eye. This means that eyes contain clear, high contrast edges, which can be easily detected by our visual system (Langton, Watt, & Bruce, 2000). Although another's gaze is used to communicate signals such as threat in many animal species, the ability of humans to signal more complex and nuanced messages through gaze appears to be an evolutionary trait which separates us from the animal kingdom.

1.2 Relevance of gaze to indicate something of interest

One example of a complex gaze processing effect is reflexive gaze cueing. This occurs when a person is presented with a face looking either to the left or right, and their attention is shifted automatically to the direction indicated by the gaze. This was shown by Driver et al (1999) in a study where observers fixated a central point, where a face would appear looking rightwards or leftwards. Observers then had to say what letter was subsequently presented on one side in their periphery (similar to previous attention cueing paradigms using arrows instead of eyes e.g. Posner, Snyder and Davidson, (1980). Observers were always faster to judge letters that appeared on the side cued by the initial gaze direction. This held whether they were told to ignore the gaze direction and even when the gaze was consistently predicting the wrong direction for the letter. This effect has also been demonstrated even when a very simple cartoon face is used for the cueing (Friesen & Kingstone, 1998). It has been suggested that gaze cueing effects are exogenous and automatic, having developed during evolution as a mechanism to signal threat (Emery, 2000). We attend to what another person is attending as early as 3 months old (Hood, Willen, & Driver, 1998), possibly because it is likely to signal threat or something else of interest. There is some evidence that suggests that gaze cueing effects are in fact endogenous and possibly controlled by frontal brain regions. For example, a patient with damage to their frontal lobe did not produce gaze cueing effects, even though their peripheral attention cueing was normal for other tasks

(Vecera & Rizzo, 2006). In either case, it is clear that gaze direction has a powerful effect on where our attention is directed.

1.3 Gaze to indicate interest in another person

A very significant part of social gaze perception is judging direct gaze, that is, gaze that is directed towards the viewer. In the animal kingdom this direct gaze is used as a symbol of threat and dominance (Emery, 2000). In humans, direct gaze can influence a variety of social tasks and captures attention more than averted gaze (Senju & Hasegawa, 2005). For example, Adams & Kleck (2005) have shown that the gaze direction of an actor (either direct or averted) can influence his/her perceived emotion expressions. They found that an actor with direct gaze was more likely to be perceived as angry or happy (emotions associated with approach behaviours) whereas they were perceived as fearful or sad with averted gaze (associated with avoidance behaviours). It has also been shown that direct gaze enhances face recognition in both adults (Hood, Macrae, Cole-Davies, & Dias, 2003; Vuilleumier, George, Lister, Armony, & Driver, 2005) and young infants (Farroni, Massaccesi, Menon, & Johnson, 2007) and that the feeling of being looked at can vary with both stress levels (Rimmele & Lobmaier, 2012) and social anxiety (Jun, Mareschal, Clifford, & Dadds, 2013). Together these results demonstrate that direct gaze exerts a powerful effect on social perception.

1.4 Gaze perception in abnormal development

Normal gaze perception, that is, the ability to quickly and accurately perceive the gaze deviation of another person, is a critical part of human development and has been shown to be abnormal in both autism spectrum disorders (ASD) and schizophrenia. Typically developing children will attend to an adult's gaze within days of birth (Farroni et al., 2002) and will engage in triadic attention, where the child will follow an adult's gaze to another object, forming a triad between the child, the adult, and the object (D'Entremont, Hains, & Muir, 1997). This ability to engage in joint attention with another person is a critical part of Baron-Cohen's theory of mind, a highly influential theory on the causes of autism (Baron-Cohen et al., 2001, 2001). Baron-Cohen suggests that a crucial part of social interaction is interpreting the mental states of another person, which requires an understanding of where another person's attention is directed. He found that individuals with autism were less able to recognise the mental states of people from images of their eye region. Since this pioneering work a large amount of autism research has focused on understanding how those with ASD perceive another's gaze.

Research into ASD has shown that those with autism fixate on the eye region of faces far less than those with typical development; instead choosing to fixate the area around the mouth (Neumann, Spezio, Piven, & Adolphs, 2006). Further work has shown that it is not simply that those with ASD fail to fixate the eye region, but that they will actively make eye movements away from the eye region if they are initially fixating it (Kliemann, Dziobek, Hatri, Steimke, & Heekeren, 2010). Similarly, when forced to fixate the eye region for a task, those with ASD perform worse than typically developing children on gaze following tasks (Leekam, Hunnisett, & Moore, 1998) and produce abnormal EEG responses when observing gaze stimuli (Grice et al., 2005; Senju, Tojo, Yaguchi, & Hasegawa, 2005). Specifically, participants with ASD had a weaker EEG response to direct gaze than age matched controls; a pattern that is consistent with infant gaze processing, suggesting deficits in ASD may be due to delayed development of gaze processing mechanisms (Grice et al., 2005).

Similar deficits have been observed in people with schizophrenia. It has been suggested that some of the paranoid symptoms associated with schizophrenia could be the result of impaired gaze judgements causing those with the disorder to believe they are being looked at when they are not (Rosse, Kendrick, Wyatt, Isaac, & Deutsch, 1994). Interestingly, research has shown that those with schizophrenia are able to discriminate between leftwards and rightwards directed gaze as well as control subjects (Franck et al., 1998), however, recently a separate study has shown that if the task is to discriminate between direct and averted gaze, schizophrenics perform worse than typical controls (Franck et al., 2002).

Taken together, the evidence in this section emphasises the importance of gaze processing for social cognition. Gaze cues have evolved to be highly salient and can induce shifts in our attention. It is important to understand how gaze perception functions under a wide range of conditions so that we can understand why we are so sensitive to it and why it seems to fail in ASD and schizophrenia.

2. How do we make judgements about the direction of social attention?

In the previous section, the numerous reasons for studying how social cues are processed were established. Here, we will discuss research on the mechanisms and limits of how humans process social cues from gaze and head direction; from early work establishing the precision of gaze deviation detection, to recent modelling and imaging research.

Early research established that acuity for gaze direction was very good (Gibson & Pick, 1963). The researchers used a design where a demonstrator was required to fixate on a target behind the participant, who then had to judge the direction the demonstrator was looking in. They found that people were accurate to approximately one minute of an arc (1/60th of a degree). This gave an early indication of how precisely the brain can represent gaze direction, suggesting there may be neural processing dedicated specifically to it.

As mentioned earlier, the human eye contains high contrast edges between iris and sclera which could potentially be used to judge eye gaze position, possibly by calculating the distance between the edge of the iris and sclera and the corner of the eye. Although this would provide a simple and theoretically effective mechanism for judging gaze direction, there is evidence that suggests this is not the case. Ricciardelli, Bayliss and Driver (2000) found that gaze direction was not simply derived from the luminance profile of the eye (e.g. simply position) by demonstrating that when the luminance polarity of the eye region was reversed, the processing of gaze was significantly disrupted. For example, observers were more accurate in judging gaze direction in figure 2a than 2b. This suggests there must be a more specialised mechanism for processing the eye region that is separate from the processing of low level visual features, such as contrast.

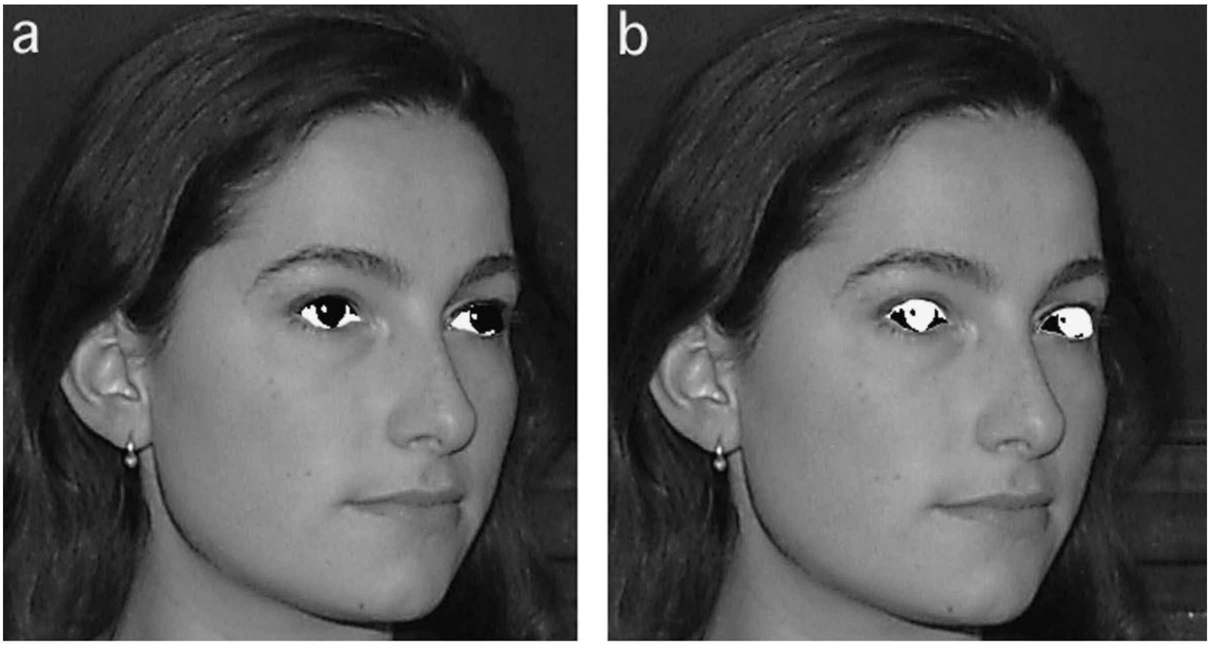

Figure 1:2 - Example Stimuli taken from Ricciardelli, Bayliss and Driver (2000)

. (a) A gaze direction stimulus with two tone eyes with regular polarity. (b) As in (a) but with the polarity of the eye region reversed. Observers more accurately judge gaze direction for (a) compared to (b).

2.1 Neural encoding of Gaze Direction

Adaptation techniques have been used to gain further understanding of the mechanism by which gaze direction is encoded, specifically, whether gaze is encoded by an opponent or multi-channel system, both of which exist in the human visual system. Adaptation paradigms exploit the fact that when an observer is exposed to a specific stimulus for a period of time, the cells which process that stimulus become less sensitive, which produces biases in the perception of subsequently presented stimuli. For example, if an observer is shown a red patch for a few seconds and the patch is then replaced by a grey screen, the observer will see a green patch. This occurs because colour is believed to be processed in an opponent manner and their red sensitive cells become fatigued and only the green sensitive ones will fire, producing the colour after effect.

An opponent channel system functions by having two pools of neurons, which are sensitive to opposite ends of a representation axis. For example, in the colour processing system, there are neurons in the lateral geniculate nucleus (LGN) which are sensitive to either red or green light. By combining the outputs of these cells, any colour on the axis between red and green can be represented (Webster, 1996). A mechanism such as this could plausibly encode gaze direction, where the two opponent channels code for leftwards and rightwards gaze and their relative outputs can be combined to form any point between the extremes. In a system like this, adapting one of the two opponent channels will bias the perception of subsequent stimuli toward the non-adapted end of the axis. An alternative mechanism is a multichannel system, where there are a number of separate pools of neurons, each sensitive to a particular direction. This is the case for orientation; there are many neural channels in the visual system that code different orientations. Each channel responds maximally when an edge is presented which matches its preferred orientation and reduces in its response as the orientation moves from this peak orientation (Hubel & Wiesel, 1968). By comparing outputs from the all responding channels, the visual system can determine any orientation presented. In this case, adapting to a given orientation reduces the response of neurons to that particular orientation, so subsequent testing of orientations that are not close to the adaptor will not be affected.

Calder et al**.** (2008) showed that gaze direction is not simply processed by a two channel (leftwards/rightwards) opponent system (fig 3a) but rather a three channel non-opponent (fig

3b) system. In a first experiment, participants adapted to direct, leftward or rightward gaze and were then tested on a range of gaze directions. Responses to each direction were reduced by the adaptation, but only when the direction that was adapted was the same as that tested. This excludes a two channel system, as direct gaze would activate both left and right channels equally (as it falls in the centre of the leftwards/rightwards axis), which would lead to no change in the responses when direct gaze was adapted (compare fig 3c and 3d). In a second experiment, participants adapted to interleaved leftwards and rightwards directed gaze and were then tested on a range of gaze deviations. Participants were more likely to respond that small offsets to the left and right were "direct" after this adaptation, which would not be predicted by an opponent channel model, as both left and right channels would be equally adapted, leaving direct responses unchanged. The authors interpreted their results as evidence of a multichannel system, with three channels (direct, leftwards and rightwards) that encode all directions of gaze and can be individually adapted out.

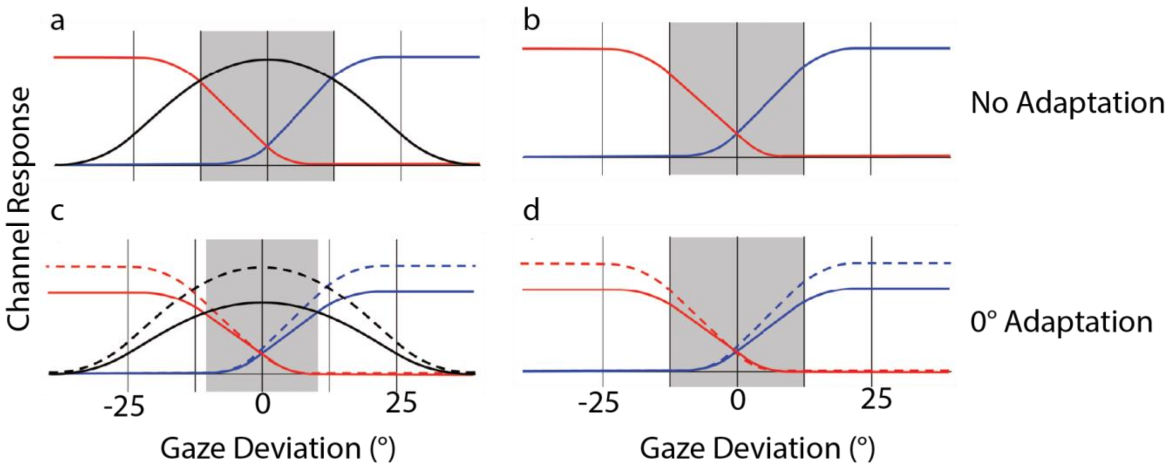

Figure 1:3 – Schematic hypothetical models for response channels processing gaze direction (adapted from Calder et al, 2008)

. A: The responses of the three gaze processing channels in a theoretical three channel, non-opponent system. Separate channels encode leftwards (red), direct (black) and rightwards (blue) gaze deviations. B: Theoretical responses of a two channel opponent model of gaze processing. Gaze deviations are encoded

as a combination of leftwards (red) and rightwards (blue) channel responses. C/D: The theoretical responses of the models in A/B after adaptation to a 0° gaze deviation stimulus (i.e. direct gaze). Solid lines show post-adaptation channel responses, dashed lines show pre-adaptation. D predicts no change in response to subsequent direct gaze stimuli whereas C predicts a reduction in direct processing, which is consistent with the behavioural data collected by Calder et al.

2.2 Neural substrates of gaze processing

The underlying neural activity that drives gaze processing has been found to be located in the superior temporal sulcus (STS), of the parietal cortex. Perrett al. (1985) used electrophysiological recordings from the brains of macaque monkeys to show the existence of cells that were selectively responsive to direct, left averted and right averted gaze within the STS. This supports the model of three non-opponent channels (Calder et al., 2008) and is consistent with research in humans; Campbell (1990) showed that those who had suffered damage to this region could no longer process eye gaze directions. A recent study (Calder et al., 2007) combined adaptation and fMRI to show the existence of different populations of neurons tuned to averted gaze directions, similar to the data reported in monkeys. Responses from different regions in the STS and inferior parietal cortex were selectively reduced by adaptation to either left or right averted gaze.

Head rotation, another important cue for the direction of social attention, is also encoded in a similar manner. In Perrett et al.'s (1985) study, cells were found in the STS which are sensitive to head rotation in macaque monkeys. Separate areas of the STS were activated by images of other monkeys either facing directly toward the subject or in profile view, facing away to the left or right. The existence of the same three-channel mechanism found for gaze processing has been reported by Lawson, Clifford and Calder (2011). They used a very similar adaptation method to Calder et al. (2008), to show that head rotation adaptation produced the same results as for gaze (i.e. reduced responses after adaptation to rightwards, leftwards and direct facing heads), for both the horizontal and vertical axis.

2.3 Cardinal Biases in Gaze perception

A simple account of how gaze direction could be processed in the brain is a mechanism that determines the overall gaze deviation by simply combining horizontal and vertical components of the eye direction. This would allow all possible gaze deviations to be represented as some combination of vertical and horizontal offsets, efficiently converting the complex representation of the eyes into a simple output that could be used in further processing. This idea of a dominance of cardinal directions is consistent with prior research into visual processing. For example in orientation perception, judgements are found to be more accurate for, and biased toward, cardinal axes (Girshick, Landy, & Simoncelli, 2011). Similar results are found for motion, where the sensitivity for detecting motion in noise is greater for cardinal directions (Morrone, Burr, Pietro, & Stefanelli, 1999). This is also consistent with the findings of Bock and Dicke (2008), who used a triadic experimental paradigm, a less common technique used for measuring gaze direction where participants must judge what object a “sender” (or gazer) is looking at. They found that participants are less accurate and biased toward cardinal directions when they have to judge “sender” directions that were on diagonal axis. This supports the idea that eye direction processing either uses cardinal axes as the main cue for gaze direction or has non-cardinal axes that are less sensitive.

Evidence that at least one non-cardinal gaze mechanism exists has been recently shown by Cheleski et al. (Cheleski, Mareschal, Calder, & Clifford, 2013) using an adaptation paradigm. Previous research has shown that adaptation after-effects can be produced from gaze stimuli, with observers being less sensitive to leftwards or rightwards gaze after long exposure to that direction (Jenkins, Beaver, & Calder, 2006). Notably, these adaptation effects persisted across faces of different sizes, different identities and different head orientations, suggesting the effect is not the result of low level adaptation to position.

In Cheleski et al.'s study, participants first adapted to interleaved gaze directions that were either diagonally up-right/left-down or left-up/right-down. They were then tested on the same (congruent) and opposite (incongruent) direction. Results suggested that adaptation had larger effects after congruent adaptation than incongruent. This provides strong evidence for gaze processing to not be exclusively encoded by cardinal mechanisms. If this were the case, both adaptation conditions would activate up/down and left/right mechanisms equally and there would be no difference between congruent and incongruent adaptation.

2.4 Role of direct gaze: Categorization experiments

It has been established that direct gaze is a very important stimulus since it influences emotional processing (Adams Jr. & Kleck, 2005), there are specific cells selective for direct gaze (Perrett et al., 1985), and recent models of gaze processing suggest there is a channel of processing dedicated to direct gaze (Calder et al., 2008). This has led to research designed to understand what direct gaze is and what makes it special. Gamer and Hecht (2007) discovered that there is a fairly broad range of gaze directions that an individual will perceive as directed at them, which they called the “cone of direct gaze” (CoDG). They used a method of limits technique, where participants moved sliders that changed the direction of gaze of a stimulus face. The face either started with the eyes converging directly on the observer or pointing far from the observer. In either case, the observer adjusted the eyes until they reached a point where they no longer judged the gaze as direct (if the eyes started on the observer) or when the observer felt the face was now looking at them (if the eyes started away from the observer). The points at which this occurred (to the left and to the right) defined the edges of the cone of gaze directions where the observer felt they were being looked at, which they found to have a quite wide range of 8^{o}-9^{o}; wider than would be necessary due to limits in an observer’s acuity.

Subsequent research into the CoDG has employed a new technique, developed to remove issues associated with the traditional method of limits, such as participants’ tendency to not change their decision (e.g. from direct to not direct) until they are certain a change has occurred, possibly ignoring points where subjectively it had (Hock & Schöner, 2010). The new method requires participants to categorise the gaze in computer generated faces as either left, right or directed at them over a range of randomised gaze directions. The proportion of times an observer responded left, right or direct for each direction is plotted, and logistic functions fitted to the leftwards and rightwards data. A function is then fitted for the proportion of direct responses as one minus the sum of the left and right responses. The CoDG is then extracted from the points where the left and right lines intersect the direct line (fig. 1) (Ewbank, Jennings, & Calder, 2009).

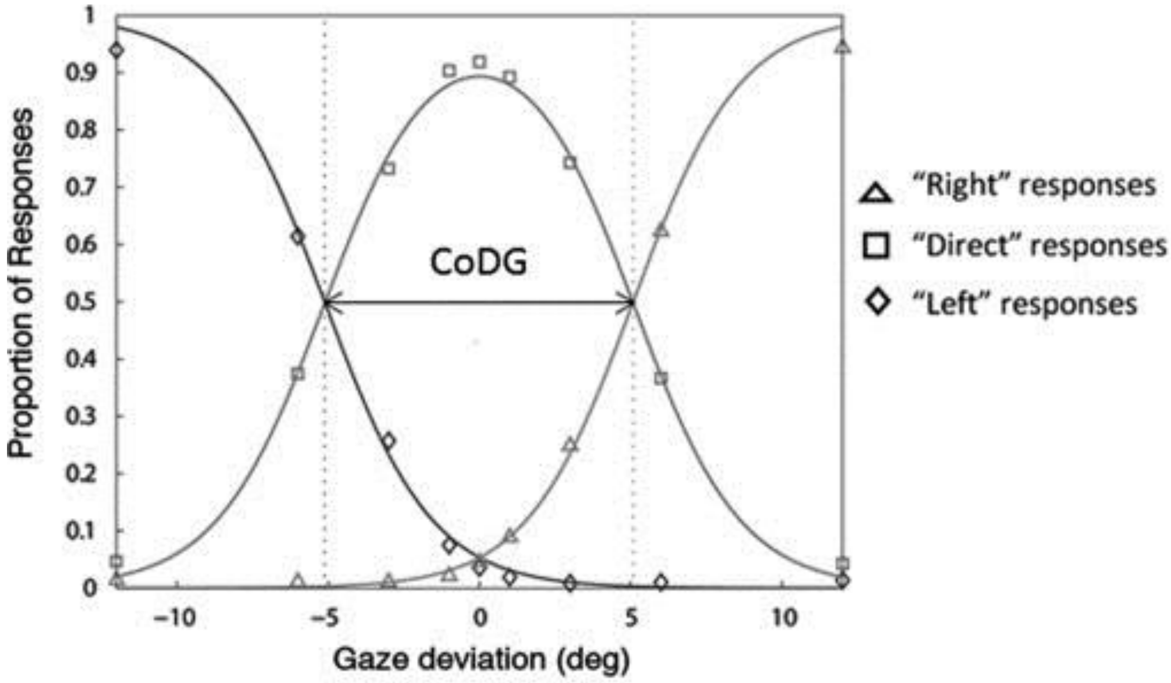

Figure 1:4 - Example of a measured Cone of direct gaze, adapted from Jun et al (2013)

. Proportion of responses "right" (triangles) and "left" (diamonds) with logistic functions fit to the points. Data for "direct" responses (squares) and function fit, calculated as one minus the sum of the logistic functions for left and right. The dashed lines show the range of the cone of direct gaze.

The new CoDG analysis method has been used to show that the CoDG can vary between individuals and under different conditions. Ewbank et al. (2009) found that the cone width increased when people were viewing an angry face, whereas neutral and fearful faces showed no difference. Vida and Maurer (2012) found that the CoDG was larger in children aged 6 compared to adults, suggesting that they were more biased to perceive another person as looking at them. More recently, Jun et al. (2013) found that the cone was larger in male individuals with anxiety disorder, showing that it can vary with personality traits.

2.5 A Prior for direct gaze

Although the CoDG provides a useful tool for understanding differences between individuals' processing of direct gaze, it does not provide information about the specific mechanisms that underlie these differences. The understanding of our processing of direct gaze took a step further by the discovery that people have a prior expectation that gaze is directed towards them. Mareschal, Calder and Clifford (2013) found that when luminance noise was added to the eye region of faces (increasing the uncertainty in the stimulus), the probability that participants would categorise eye directions as "direct" (rather than leftwards or rightwards) increased. They applied a Bayesian framework, where a decision about gaze made under uncertainty (e.g. when noise is added) is influenced by a prior expectation of the probability of an event happening. This revealed that in humans, this prior expectation is that gaze is directed towards them. They also show that it is not simply the case that noise from the eye region causes participants to respond to the head direction, which was directly facing the participant. This was shown by repeating their experiment but using heads that were either rotated leftwards or rightwards. They still found evidence for a prior for direct gaze; although they also reported that head orientation influenced the overall perception. A further study confirmed that the prior for direct gaze influenced gaze perception in all directions, not just a horizontal axis (Mareschal, Otsuka, & Clifford, 2014).

3. How do we integrate gaze and head direction into a single percept?

Although a great deal of research has focussed on the processing of the eye region in relation to gaze direction, it is important to note that it is not only information from the eyes that determines the perceived direction of gaze. Wollaston (1824) first demonstrated the influence that head rotation has on perceived gaze direction, by designing faces where the eyes were identical but the surrounding features were facing in different directions. This results in two faces that appear to be looking in entirely different directions (fig 5) even though the eye regions are identical. Langton and Jenkins (2003) further highlighted the relationship between gaze and head rotation by showing that the configuration of the entire head was key to processing gaze direction. They created stimuli where the configuration of the face was disrupted by either rotating the eyes 180^{o} in the face or rotating the entire head by 180^{o} but leaving the eyes the correct way up. They then tested the gaze direction discrimination threshold of participants. Their results showed that both disruptions to the configuration of the face increased the discrimination threshold of gaze direction. This demonstrates, in line with the above finding, that gaze processing is influenced by the head direction rather than being an exclusively eye dependent process.

Figure 1:5 – An example of the Wollaston illusion

. In a and b the eye region is identical, the only difference is the surrounding facial features, which creates the perception that the faces are looking in different directions. Figure adapted from Wollaston (1824).

Further research has expanded on the importance of considering the combination of eye and head cues in gaze processing. Todorovic (2009) was able to show that even in a basic schematic face; the eccentricity of the features of the face (for example the position and configuration of the nose) can influence perceived gaze direction. The authors held the iris eccentricity (i.e. the direction the eyes are pointing) constant and changed the eccentricity of the features within the face (fig 6). When observers judge the gaze deviation of these faces, the perceived direction of gaze was biased towards the direction of the face. In this case, the head exerts an attractive effect on gaze direction.

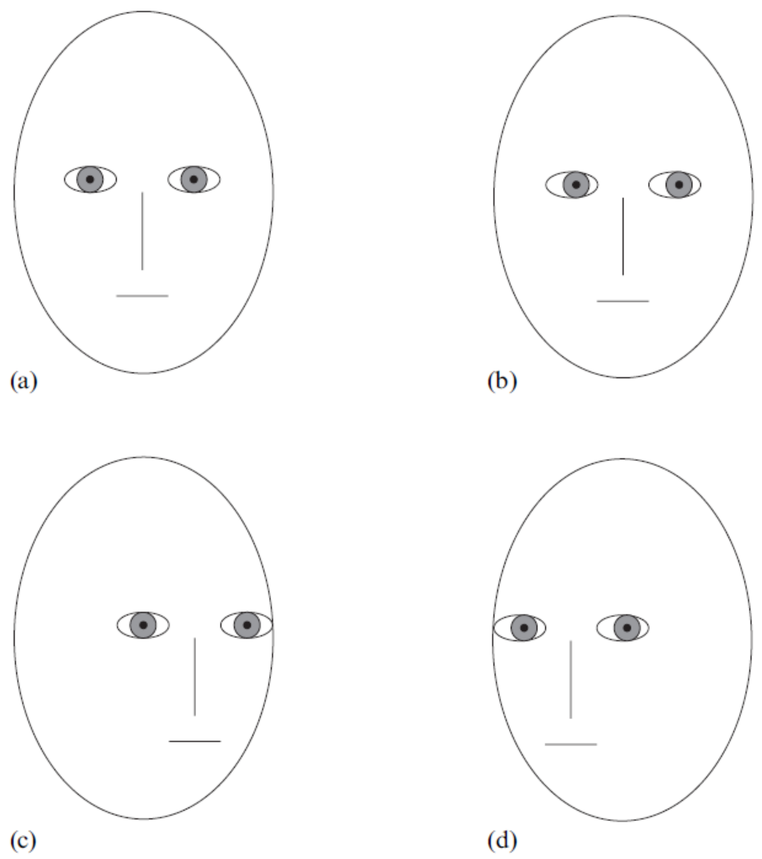

Figure 1:6 - Example schematic face stimuli taken from Todorovic (2009)

. Faces a/c and b/d have identical gaze directions but shifted face eccentricities. The perceived gaze direction of c and d is shifted in the direction of the face eccentricity.

The effect of head turn on perceived gaze direction is not consistent solely with an attractive effect (where the direction of gaze is perceived to be pulled towards the direction of the head), as has been reported in some cases (Langton, Watt, & Bruce, 2000). Nor is it solely a repulsive effect, where the direction of gaze is perceived to be away from the direction of the head, (Anstis, Mayhew, & Morley, 1969). Ricciardelli and Driver (2008) suggest that the discrepancies observed in the literature are due to the time pressures put upon the gaze direction decision. They found that when participants were told to respond as quickly as possible, a congruent (attractive) effect was found, where participants were more likely to categorise the gaze as pointing in the same direction that the head was facing. When no time pressure was included in the instructions, the opposite, incongruent (repulsive) effect was observed.

An alternative explanation for the conflicting evidence for attractive and repulsive influences of head orientation on perceived gaze direction has been suggested by Otsuka et al. (2014). They examined how head and eye information was combined, using information from the eye region only (e.g. cropped images not showing the full head), or the whole head. They measured gaze categorization with these two types of stimuli and found greater repulsive effects using the cropped heads than with the full head conditions. They suggested that the rotation of the eye region within a rotated head was the factor that caused a repulsive effect from head direction, as the visible amounts of sclera change as the head moves round. Since repulsion was weaker in the whole head stimuli, they proposed the existence of a (weaker) attractive effect that was due to the global head direction. Using a linear regression model, they were able to derive weightings of the head and eye information and show that head rotation has a negative (i.e. repulsive) weighting on the eye-region information but a positive (attractive) weighting on overall perceived gaze direction.

4. Perceiving social cues in the periphery

Most research into the processing of social cues has focused on how we judge gaze deviations that are presented in central vision. This thesis aims to extend these findings to beyond this limited, optimal viewing condition. One question which is not well understood is how we perceive social cues in our visual periphery. Visual processing away from fixation is limited by lower visual acuity and crowding, greatly reducing peoples' performance on even simple visual tasks. In this section these limitations on the processing of peripheral stimuli will be discussed.

4.1 Limits in peripheral vision: resolution

Visual perception has been shown to be limited across a large number of simple visual tasks. Work from Rovamo, Virsu and Hyvarinen (1982) showed that observers' ability to detect the presence of a fixed contrast black and white grating stimulus decreased rapidly as the position of the grating moved away from fixation. Similar results have been found for both reading and numeral perception. Chung, Mansfield and Legge (1998) found that participants' maximum reading speed was significantly lower in the periphery, even when very large letters were used and Nasenen and O'Leary (1998) showed that perception of hand written numerals was much weaker in the periphery. Clearly peripheral vision is worse than foveal vision; however it serves an important purpose by guiding future shifts in attention. It is therefore important to better understand how gaze may be processed in the periphery.

In order to understand how complex stimuli are processed in the periphery, the causes of limited processing must be established. The first of these is reduced peripheral acuity. The human visual system is designed, such that the central area around where we are fixating is processed in very fine detail and the representation becomes rapidly less precise as you move away from the centre (e.g. fig 7). This is a result of the biology of the eye and the proceeding layers of visual system. The eye focuses the visual scene onto the retina, which contains a distributed array of light sensitive cells, known as rods and cones. The number of cells available to process an area of a visual scene sets the limit on the precision with which that area can be represented, the larger the number of cells, the more precise the representation. At the centre of fixation (the fovea), the density of these light sensitive cells is much higher than the periphery (Curcio & Allen, 1990).

Figure 1:7 – Example of how our peripheral vision quickly reduces in resolution

. Left shows the real image and right shows how out visual system would process it if we were fixating in the centre. http://anstislab.ucsd.edu/2012/11/20/peripheral-acuity/

The increasing size of the cortical region designated to process a given area of the visual field from the periphery to the fovea is known as cortical magnification (Duncan & Boynton, 2003). In order to match the perception of peripheral objects to those presented in the fovea, a technique known as M-scaling has been developed. Objects are increased in size when they are presented in the periphery to match performance to that in the fovea. This has been done successfully with grating detection, orientation and motion (Rovamo & Virsu, 1979.; Virsu & Rovamo, 1979). The exact relationship between the location in the periphery and the increase in size necessary to match the fovea is a subject of debate and a single rule that applies for all stimulus types under all conditions has not been defined. Recently Duncan and Boynton (2003) used both Vernier and grating stimuli to estimate the cortical magnification factor for human participants. They find that the scaling factor required (M) for a given eccentricity (E) can be determined using the equation: $1/M = 0.065E + 0.054$.

4.2 Crowding

There are situations where observers are still unable to complete certain tasks even using M-scaled stimuli to reduce issues of cortical magnification. This is due to *visual crowding*, where elements presented in the visual periphery cannot be distinguished individually but rather appear “cluttered”. Researchers have suggested that this process is a form of compulsory averaging, whereby similar elements in the periphery are grouped and processed as an average rather than individually (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Parkes et al. presented participants with a single oriented Gabor patch, surrounded by another 8 Gabor patches of different orientations and asked observers to report the orientation of the central grating (Fig. 8). They found that observers could not accurately report the orientation of the central grating; however, they were able to accurately report the average orientation of the group of gratings. This demonstrates that although information about the individual elements is lost, the average is preserved. Together both crowding and reduced resolution serve to limit peripheral processing.

Figure 1:8 - Example of two stimuli from a crowding experiment taken From Parkes et al. (2001)

. Above: A crowded stimuli, when fixating the cross, the orientation of the central Gabor patch cannot be determined. Below: The same central Gabor, now without the flankers. Without the flankers the orientation can be perceived more easily then fixating the cross.

4.3 Peripheral emotion and gender processing

In many situations we are presented with groups of faces simultaneously, such as in group conversations or when giving a talk or presentation. In these cases, it is necessary to be able to perceive faces in your periphery, in order to shift your attention appropriately and communicate effectively. There is already some research which addresses peripheral perception of faces for various tasks. Makela et al. (2001) found that observers' ability to determine the identity of faces was much worse in the periphery and required increases in both size and contrast to match performance at fixation.

Much of peripheral face perception research has focused on emotion perception. This is driven by the finding that emotional content has been found to draw attention and be better recognised than other, neutral stimuli in the periphery (Calvo & Lang, 2005). Bayle et al. (2011) asked participants to discriminate the gender and the emotion of faces presented at a number of visual eccentricities. They found that emotion discrimination was always better than gender in the periphery and that the difference between the two tasks increased with eccentricity. Interestingly, a recent study (Calvo, Fernández-Martín, & Nummenmaa, 2014) has shown that participants are better a recognising happy faces than either fear or surprise. This is particularly relevant to this thesis as both fear and surprise are expressed through the eye-region, whereas the most salient cue for happiness comes from the mouth. This suggests that the eye region may not be processed accurately when seen in the periphery. This may be because eyes are particularly susceptible to crowding because they are two similar objects, always presented close together.

4.4 Peripheral Gaze Perception

The issue of peripheral gaze perception has been addressed to some extent in existing literature. Loomis (2008) conducted experiments investigating the precision with which both head rotation and gaze deviation are represented in the visual periphery. To measure peripheral head rotation perception, participants were required to adjust a graspable pointer to match the head rotation of a demonstrator. They found very accurate performance, with almost no reduction in accuracy between 0 and 45 degrees of eccentricity. Even as far as 90 degrees eccentricity, observers were still responding well above chance to the presented head rotation. This clearly shows that head rotation can be represented in the visual periphery. In contrast, when observers judged the gaze deviation of a face presented on a computer screen, performance was much worse in the periphery. At just 8° eccentricity, participants' responses were tending towards direct. This suggests that at this eccentricity, participants were simply relying on the head rotation (which was always direct) to make their judgements, as their perception of the gaze direction had become too unreliable. This has interesting implications for the interaction between head and gaze deviation in the periphery, as the weighting between these two to determine overall gaze direction may change as a function of the reliability of the gaze information.

Yokoyama et al. (2014) have also investigated peripheral gaze processing. They presented observers with a central task (identifying a letter) while gaze stimuli were presented in the periphery. Participants either had to determine if the gaze of the faces was directed at them (direct) or away from them (averted), or they had to discriminate between leftwards and rightwards directed gaze. The authors found that observers were able to discriminate between direct and averted gaze, but not between the two averted gaze directions. Contrary to Loomis et al., this suggests that some gaze information is retained in the periphery. It may be that because all the faces in the study were forward facing, direct gaze perception was facilitated, whereas averted gaze in either direction was harder to resolve.

Very recent research (published after similar work in chapter 1), has provided further insight into peripheral gaze perception. Palanica and Itier (2015) presented faces which could have a combination of either direct or averted, gaze deviation and head rotation and asked participants to classify the faces as “direct” or “averted”, across a range of fixation eccentricities. They found that discrimination performance reduced significantly when faces were presented beyond 6° of eccentricity; this may be the maximum eccentricity where peripheral gaze perception is possible. Their results also showed that participants were significantly faster to respond to peripheral “direct” gaze when the head rotation was also “direct”, suggesting that head rotation plays an important role in peripheral gaze perception.

5. Perception of groups of objects: spatial averaging

The previous sections dealt exclusively with perception of gaze (or head) for one stimulus only. Here we will examine how groups of face stimuli are processed. In the majority of psychophysical research, stimuli are presented in isolation with plain backgrounds at fixation. This does not very accurately reflect the makeup of any real world scene. Research has shown that observers can still make judgements about natural scenes, even when a large amount of information is removed, by exploiting the redundancy and predictability of these scenes (Kersten, 1987). One way that humans exploit this redundancy is by representing groups of objects as ensembles and calculating summary statistics about their properties, rather than processing every element individually. For example, when we see a tree covered in leaves, we have a gist perception of the average colour, shape and size of the leaves without needing to analyse each leaf individually. This already occurs automatically under crowded conditions, as previously mentioned (Parkes et al., 2001a), but there is also strong evidence that this averaging can be done voluntarily.

5.1 Summary Statistics for visual properties

The ability to average allows us to represent a large amount of complex information in a single output, though it seems that this comes at the cost of sacrificing precise information about the individual elements. Ariely (2001) demonstrated this with a size averaging task. Participants were presented with sets of white dots of different sizes on a grey background followed by a single dot. The participant then completed one of two tasks, either membership or average size. In the membership task, the participant had to say if the single dot was presented as part of the previously shown set, and in the average task they had to say if the previously shown set was larger or smaller, on average, than the single dot. Performance was surprisingly poor in the membership task, with participants incorrectly categorising elements close to the mean as part of the set, even when not present, and rarely categorising elements far from the mean as part of the set. Averaging performance was much better, suggesting participants were maintaining a representation of the average at the expense of the individuals.

Using a similar method to Ariely (2001), Haberman and Whitney (2007, 2009) have shown that this ability to extract summary statistics for visual properties is not limited to simple dot stimuli, by showing that people can average both emotion and gender from groups of faces. In one study, they created a set of 50 faces, morphed between 100% happy and 100% sad. They then presented sets of four of these faces for two seconds, followed by a single face and required participants to judge if the single face was more or less happy then the average of the set. Like Ariely, they find accurate discrimination for the mean emotion of the set. They also find in a control condition, that participants perform at chance when asked if the subsequently presented face was part of the original set, suggesting that information about the individual items is not preserved after averaging. Similar results were found for averaging of gender. Haberman and Whitney also report that when the sets were only presented for 500ms, mean emotion discrimination was still very good, suggesting that the mechanism for averaging acts rapidly and efficiently.

As well as being able to represent the mean of a set of objects, there is also evidence that observers can perceive the variance of groups of orientation patches (Dakin & Watt, 1997; Solomon, 2010). Solomon (2010) measured observers’ performance when discriminating either the mean or the variance of the orientation of groups of Gabor patches. He found that observes could more efficiently estimate the variance of the sets than the mean. This may seem counterintuitive, as the variance is by definition the deviation from the mean. The author suggests that observers may actually be using some other statistic, such as the range of the distribution, rather than the mean, to make their judgements which might explain why they are able to judge the variance more accurately than the mean.

There is also evidence that observers can represent multiple ensembles simultaneously. Chong and Triesman (2005b) presented observers with sets of dots of different sizes (similar to Ariely, 2001) and had them complete a size averaging task. The difference here was that the dots were of two different colours and the observer had to respond to the mean of one of the sets. They found that mean size discrimination thresholds were equal whether observers were pre-cued to which colour of dots they had to respond to or were only cued after the sets were presented. These thresholds were also no worse than those for averaging a single set of dots and could not have been achieved by simply combining the mean of the two groups to make their response. The authors suggest that this is evidence that averaging can occur automatically and in parallel, without attention. Subsequent research, using a similar task with greater control over the difference between the target and distractor distributions (Oriet & Brand, 2013) suggested that although observers are able to average two sets simultaneously, this comes with some processing costs. Discrimination thresholds were higher in a post-cued task, where the observer was required to process both averages as they didn't know which would be tested, than in the pre-cued task where they only needed to attend to one set.

Whether or not there are generic mechanisms for all averaging tasks, or separate mechanisms which only integrate select information, is not fully understood at this stage. There is at least some evidence suggesting that there is not simply one averaging mechanism (Haberman, Brady, & Alvarez, 2015). These authors measured averaging performance for a large number of participants on both low level stimuli (colour/orientation) and more complex face stimuli (emotion/identity). They found that observers' performance was correlated between the two face tasks, but not between the face tasks and the low level stimuli. They suggest that there are different limits on the integration of these different types of stimuli, eliminating the possibility of a single generic integration unit.

5.2 Processing groups of social cues

Understanding how we perceive groups of faces is particularly important, as we are often required to process a group of faces together, either in a conversation or in a crowd. Information from a group of people can be more informative than from an individual, such as indicating the source of a threat or the location of an object of interest. Gallup et al. (2012), put groups of actors onto a street in various groups sizes and asked them to look in a certain direction. They then measured the number of pedestrians who looked where the actor group was looking as they passed by. They found that a larger group of actors was more likely to cause shifts in passer-by attention than a single actor or a small group. Evidently, humans are sensitive to the attention of a group of people, potentially even more so than an individual.

Particularly relevant to this thesis, is a recent study which investigated the averaging of gaze deviation (Sweeny & Whitney, 2014). The authors presented either 1, 2, 3 or 4 faces from a set of 4 and required participants to judge their mean gaze deviation by adjusting an example face after the test faces had disappeared. The stimuli in their study were simple black and white outline faces and the gaze offsets were generated by manipulating the head rotation and keeping the gaze deviation fixed. This produced perceived shifts in gaze direction for the individual faces, without actually changing the position of the iris within the sclera. They found that the variance between the observers responses and the actual mean of the set decreased as more faces were presented, suggesting that they were able to average and were using more information when it was available. Interestingly, they find that there was no improvement in performance after 3 faces, suggesting that observers could not use more than three faces in their average. They also found that participants could complete this task even when the faces were presented for only 200ms so observers would not have time to sequentially fixate each face and must have been using a rapid, gist perception of the group.

6. Limits in Spatial Averaging

6.1. Equivalent Noise Analysis

In the previous section I discussed people's abilities to extract summary statistics for properties of groups of objects. For the most part this focused on whether it is possible for people to average a particular feature, or if they could average under certain conditions. In this section I will discuss the literature that shows that although people can average, they do not use the information from an array efficiently. That is, they perform as if they are only using a sub-set of items that they are presented with when they calculate the average.

The majority of this work uses a technique known as equivalent noise analysis. This technique has previously been used to estimate the neural noise in the visual system for light detection (Barlow, 1957). The key principle was that thresholds for detecting changes in light must be perturbed by neural noise in the retina, termed "dark light". An observers' total uncertainty is a combination of the variability in the stimuli and this intrinsic noise. By increasing the noise in the stimulus, researchers were able to identify the equivalent external noise needed, to match the internal noise already in the system.

In the context of voluntary averaging (as opposed to crowding), equivalent noise is used to estimate two parameters that limit an observer's averaging performance; *internal noise* and *effective sample size*. Internal noise refers to the uncertainty with which an observer judges any individual element in an array. The model assumes that all sensory representations are noisy, and that the noise is Gaussian in nature. This means that the representation of each individual element can be estimated as a Gaussian distribution, whose mean is the veridical value of the element property and whose standard deviation is the amount of internal noise associated with processing the property.

The effective sample size (ESS) refers to the number of elements that, across all trials, an observer performs as if they are combining into an average. For example, an ideal observer will always combine all the elements presented into the average calculation. If presented with 100 circles of different sizes, the ideal observer will take the linear average of all 100 and report that mean. In reality, observers are unable to integrate all the items they are presented with and tend to use only a sub-set when estimating their average. By comparing the accuracy with which an observer judges the average of a set to that of an ideal observer, the number of samples the observer must be using can be estimated (assuming the same external noise).

The two limiting factors in equivalent noise analysis, internal noise and ESS, can be estimated by measuring how averaging performance changes as a function of increasing external noise. In the context of averaging, the external noise of a set of items can be taken as the standard deviation of the items from the mean of the set. For example, in a size averaging task, if the external noise is 0, all the dots will be exactly the same size. Increasing the amount of variance in the dot sizes increases the overall external noise. When the external noise is low, an observers' ability to determine the average is only limited by their internal noise; determining precisely the size of any individual dot will provide an accurate response regardless of which dot is used. When the variance is high, observers' performance will now be limited by the number of elements they use in their average; the more dot estimates they combine the more accurate they will be. By measuring observer's discrimination thresholds (a measure of the total noise in the processing of the stimuli), at a range of external noise levels (i.e. from very low variance to very high), these two limits can be estimated. An equivalent noise function is fit to thresholds measured at each external noise level using equation 1, where σ_{obs} is the observer's discrimination threshold, σ_{int}^2 their internal noise, σ_{ext}^2 the added external noise and n_{samp} the effective number of samples used to estimate the mean (fig 4).

Equation 1: $\sigma_{obs}^2 = \frac{\sigma_{int}^2 + \sigma_{ext}^2}{n_{samp}}$

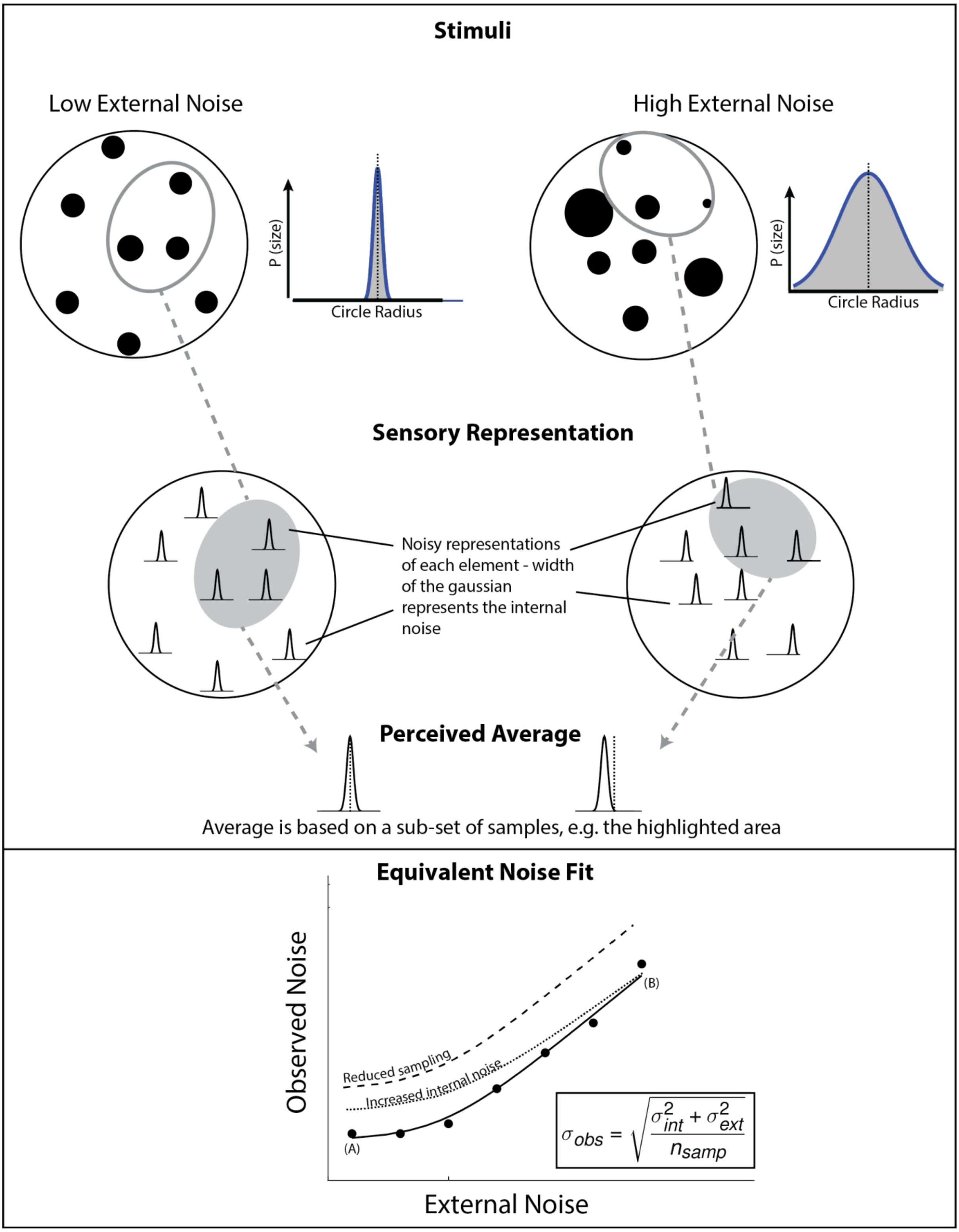

Figure 1:9- An example of the equivalent noise method for measuring the limits in averaging performance

. Stimuli are drawn from a normal distribution (e.g. the size of the black circles) and randomly positioned in an array. When external noise is high, the standard deviation of the distribution is high and vice versa

for low noise. The estimates of each element are represented as Gaussian distributions with a mean of the true value and a standard deviation that is the observer's internal noise. Observers average a sub-sample of these elements to produce their response (e.g. the highlighted areas). The observed noise (e.g. the observer's discrimination threshold) is plotted against the amount of external noise added. A function is fit using the equation in the figure and the internal noise and effective sample size are estimated.

Dakin (2001) used the equivalent noise method to investigate how these two limits affect orientation averaging under a number of conditions. He examined how these two parameters were affected by the density, size and number of elements using an orientation averaging task. Three conditions were tested, where one of these variables was held constant and the other two changed. For example, in the fixed size condition, oriented Gabor patches are randomly distributed within a fixed radius from the centre of the screen. In this case both number and density increase as more orientation elements are added to the stimulus. Results for the ESS suggest that this limit is primarily dependent on the number of elements present in the stimuli; the more elements present, the more elements observers used in their average. When the number of elements was held constant there was very little change in ESS regardless of the density of the elements. This and other similar studies, have suggested that participants will use approximately $\sqrt{N}$ of N samples. All three variables appeared to have some effect on internal noise, with density having the largest effect. The author suggests that this may be a result of densely packed orientation signals becoming crowded, degrading the representation of each individual element.

Since this technique was developed for orientation averaging it has been used to assess global and local limits in other feature domains. Research from Dakin, Mareschal and Bex (2005) used equivalent noise analysis to investigate the limits on motion direction integration. They find similar results to those for orientation, where internal noise and ESS were most affected by the total number of elements present. Using similar stimuli, Manning (2014) showed that motion averaging improved from 5-11 years and that this improvement was the result of improved element integration not any reduction in internal noise. Notably in the Manning et al. study, estimates of the ESS for participants were very low. Participants performed as if they were using as few as 1 sample from a set of 100. This is well below the established √N samples used in previous studies, suggesting that this rule may not be applicable in general and that observers can perform very poorly on these averaging tasks.

6.2 Issues with Equivalent Noise Analysis

One potential issue with the equivalent noise method is that it assumes an ideal observer will perform a linear average of the sub-set of estimates. This has been the assumption of much of the averaging literature, though there is some evidence that observers may not necessarily be doing this. Myczek and Simons (2008) replicated Ariely's (2001) study on size averaging and carried out simulations based on data from that original study and other size averaging tasks (Chong & Treisman, 2003;2005). Using an ideal observer analysis, they show that in all these cases, observers' performance could be predicted from an ideal observer who only used the maximum and minimum of a set rather than performing any sort of rapid, linear average. They suggest that the fast, voluntary mechanism proposed by previous ensemble coding research does not exist and that observers simply use a serial, focused analysis of the elements they are presented. This is consistent with situations where equivalent noise analysis has produced estimates of two or less (e.g. Solomon 2010, Manning 2014), though this explanation cannot account for cases where ESS estimates have been as high as ~40 samples in the case of Dakin (2001).

Another criticism of this method is that observers may change their strategy between high and low noise conditions. The method assumes that observers will take an average of estimates even when the external noise is very low and the task is very easy. This would predict that observers' performance should improve if they are estimating the average of four identical samples compared to a single sample as they can take advantage of redundancy in the stimulus to mitigate the internal noise associated with each sample. This was found to not be the case by Allard and Cavanagh (2012) using orientation stimuli. This presents a problem for equivalent noise analysis as it requires the same computation across all noise levels in order to estimate the two limiting parameters (internal noise and ESS). One possible resolution to this issue would be that the integration computation comes at some cost to precision. This could explain Allard and Cavanagh's findings as participants could still be averaging at low noise but this computational noise limits their performance to be similar to that of the internal noise associated with a single element. Solomon (2010) has proposed a similar model to equivalent noise to explain orientation averaging performance, which would be consistent with these multiple sources of noise. His model has the same ESS parameter as the equivalent noise model but has two noise parameters, one early, which acts at the level of the individual stimuli and one late that acts on the average calculation. This two stage noise model may be a better way to quantify the limits on averaging performance at low external noise levels.

Equivalent noise analysis assumes that all elements that are pooled together are weighted equally; however there is some evidence that this is not always the case. In fact, Alvarez (2011) used simulations to show that a weighted average, where elements which were represented more precisely are weighted more heavily than those with more uncertainty, would produce better performance than a flat weighted average. De Gardelle and Summerfield (2011) have proposed the concept of robust averaging. In this account, observers are biased to up-weight elements whose values are close to the mean of the set, and down-weight those that are far from the mean. In their study they asked participants to say if the average colour of a set of ten patches was more red or blue. By ranking the colour values presented from most red to most blue and performing a regression analysis on the ranked order of samples, they found that observers were biased to base their decision on the elements close to the mean of the group they were presented with. This does seem to suggest that observers do not weight all items equally when performing an averaging task, though this is not necessarily inconsistent with findings using equivalent noise analysis. Indeed, in order to up-weight items closer to the mean, the observer must have some perceptual estimate of what the mean is. It is possible that they do up-weight certain elements when they are making a decision using summary statistics but weight elements equally when processing the initial perceptual average. Also, equivalent noise analysis makes no assumptions about which samples an observer uses in their sub-set, so it is possible there is some bias which causes them to be more likely to include elements close to the mean in their sub-set, which could produce the same pattern of results as produced by de Gardelle and Summerfield.

Finally, although equivalent noise analysis can be informative as to *how many* samples an observer is using, it does not tell us *which* elements are being used, or if participants are biased to use certain areas of a group (e.g. the centre). A possible solution to this issue is to apply reverse correlation analysis to data collected using EQN. Reverse correlation is a method that can reveal which parts of an image an observer is using to make a decision. Usually this is done by having participants complete a task while random noise patterns are added to the test stimulus. By combining the noise patterns on trials where the participant gave correct responses, the sections of the image that the participant used to make their decisions can be determined. This technique has previously been used to reveal the tuning of depth processing filters (Neri, Parker, & Blakemore, 1999) and to generate classification images for rapid the discrimination of orientation stimuli (Mareschal, Dakin, & Bex, 2006). In this thesis, reverse correlation analysis will be applied to averaging data to reveal any biases or strategies observers use when averaging groups of faces.

7. Perception of groups of objects in time: sequential integration

Previous sections discussed how information is integrated rapidly over space by taking advantage of redundancy in visual scenes to quickly perceive global properties of groups of objects in the environment. Given that the real world is rarely static, and even when it is, our visual input can change dynamically as we make multiple saccades to points of the same object, extracting summary statistics over time is also critical to perception. For example, a single face will change expression throughout a discussion and we must make an overall judgement of mood or emotion. Here the evidence that such a mechanism for averaging over time exists, and what may limit such a mechanism will be discussed.

7.1 Limits in Sequential Integration

Observers' ability to extract summary statistics over time has only been measured relatively recently. Haberman, Harp and Whitney (2009) showed that observers could average facial expression over a set of sequentially presented faces of varying emotion, using a similar method as for spatial averaging. Observers were shown a sequence of 4, 12 or 20 faces and then had to adjust a test face match their perceived mean of the set. In a separate condition, after the set of faces was presented they were presented with a test face and had to say if it was included in the sequence they were shown. Like the results for spatial emotion averaging, they find that observers were able to accurately perceive the mean expression of the sets of faces but were near chance for their memory of individual items. As with spatial averaging, this implies that people are representing these sequences in summary form and discarding information about the individual items.

Further work has shown that the ability to represent information in summary statistics is not limited to the visual domain. Albrecht, Scholl and Chun (2012) found that observers were able to perceive the average pitch of a sequences of tones. Observers heard a sequence of tones and then adjusted the pitch of a "test" tone until it matched their perceived average. They found that observers were more accurate in the pitch averaging task than an analogous size averaging task. It is perhaps unsurprising that sensitivity to auditory sequences is greater than to those in the visual domain as auditory information is more often temporally distributed than visual information. They also show that observers could extract the mean size of circles and mean pitch of tones at the same time with minimal cost to each stimulus modality. Similar results were found by Piazza et al. (2013), who showed that, as with other ensemble coding paradigms, information about individual items is lost in favour of an ensemble representation of a set of tones. They suggest that observers are able to combine information from at least 4 out of 6 elements in a sequence of tones.

One study so far has attempted to use an equivalent noise procedure to investigate the limits of temporal averaging and to compare performance to spatial averaging using the same stimuli (Gorea, Belkoura, & Solomon, 2014). This study compared performance for temporal and spatial averaging of the size of sets of circles presented either over space or time, for different stimulus durations. The model applied was similar, though not identical to the equivalent noise model used in the spatial studies (Dakin & Watt, 1997; Dakin, 2001; Dakin et al., 2005; Manning et al., 2014). Gorea et al. (2014) use a model which includes two noise components, rather than the single "internal noise" parameter. The two noise components are separated into "early" noise which is applied to add variance to the perception of the individual elements in the set and a second, "late" noise, which is applied to the average computation between the items. Using a model with these two noise parameters and a third, ESS parameter, the authors show that participants ESSs for temporal and spatial averaging were similar (approximately 4 from a set of 8 samples). They suggest that the visual system integrates four samples with replacement, while stimuli are present, so they continuously update their estimate as more information is provided but discard the early information. This result suggests that temporal and spatial averaging may share a common cognitive limit that applies to all averaging tasks.

7.2 Mechanisms for Sequential Integration

The actual mechanism by which temporal information is integrated has been studied to some extent, with the majority of research focusing on how observers weight all the elements in a sequence. An ideal observer would weight them all equally, as is the case for spatial averaging, however there is evidence that human observers do not behave like this (e.g. the sub sampling results found by Gorea et al., 2014). Juni et al. (2012) have previously shown that the weighting of each item in a sequence can vary depending on the reliability of the information. They varied how reliably each element in a sequence of position stimuli would predict the true mean location, either so that it increased across the sequence or decreased across the sequence. Regression analysis of their responses and the locations presented suggested that, over multiple sessions, observers would increase their weightings of the more reliable stimuli and reduce those for unreliable stimuli, regardless of whether the reliable elements were at the start or end of the sequence. This goes some way to explain non-ideal performance of human observers as they are not equally weighting all items in a sequence, and shows that the temporal integration mechanism can be adapted to different stimulus properties.

Hubert-Wallander and Boynton (2015) asked participants to average sequences of 8 stimuli which could vary in size, location, motion direction and facial expression. A regression analysis was used on this data to find out how observers weight the individual items in the sequence for different stimulus types. They found that the pattern of weightings for the individual items was dependent on the type of stimulus presented. For size, motion and facial expression, they found that elements towards the end of an eight or ten element sequence were weighted more highly in their average calculation than earlier elements. This finding is known as a recency effect, where more recently viewed items are weighted more heavily in the mean calculation. In contrast, position averaging produced the opposite effect, where elements at the start of the sequence counted more towards the average than those at the end, known as a primacy effect. This suggests there isn't a single generic mechanism for all temporal averaging stimuli. It is more likely, as has been shown for spatial averaging (Haberman et al., 2015)**,** that there are separate mechanisms that respond to different stimuli types.

A possible mechanism for temporal integration has been proposed by Cheadle et al. (2014), based on previous research into robust averaging (e.g. de Gardelle & Summerfield). They suggest that an efficient mechanism for temporal integration should give more weight to signals that are consistent with what has previously been seen and that unexpected information should be ignored. They suggest this could be achieved by adjusting the gain on incoming signals, depending on how close they are to the existing mean of the sequence. They found that this model was a good fit to human data on an orientation averaging task. This model also produces a recency effect, which is consistent with previous temporal averaging data (Gorea et al., 2014; Hubert-Wallander & Boynton, 2015). This is a potentially plausible mechanism for temporal averaging, though it is not consistent with the primacy effects found by Hubert-Wallander and Boynton.

Aims of the thesis

This thesis aims to investigate how we perceive social cues when we are not looking directly at a single person. In chapter one, the question of how we perceive gaze direction in the periphery will be addressed. Particularly focusing on how we perceive direct gaze and how we integrate gaze direction and head rotation. Chapter two will investigate how social cues are averaged when presented in groups over space, using equivalent noise analysis. Finally, chapters three and four will investigate how gaze direction and head rotation are averaged over time. I will compare the findings between head and gaze averaging and aim to further our understanding of the mechanism by which temporal information is integrated.

Chapter 2 Peripheral Processing of Gaze

Abstract

When looking at someone, we combine information about their head orientation and eye deviation to judge their direction of gaze. What remains unknown, however, is how these cues combine when we are not looking directly at the person, but rather using our *peripheral vision*. Given that peripheral vision helps direct future attention, understanding how we perceive other people's gaze is key to determining their future actions. To examine this we asked participants to categorise gaze direction in faces whose heads were turned in different directions, and which were viewed using either central or peripheral vision. We report that the weight given to head orientation increases in the periphery where forward facing heads were categorised as "direct" over a wider range of eye deviations than when viewed centrally. When peripheral heads were turned, the number of "direct" responses fell for all gaze deviations with no consistent shift in left/right responses towards the head rotation. For centrally presented heads, head-orientation typically repulsed the perceived direction of gaze, and our finding of no consistent shift in responses indicates that such effects are reduced in the periphery. This is not simply the result of poorer spatial resolution in the periphery, other influences, such as crowding and priors for gaze or head direction may play a role.

Introduction

Understanding where another person's gaze is directed is a crucial component of social interaction. Gaze direction can convey information about others' intentions, but can also disambiguate communication, and alter our interpretation of another's emotion (Adams Jr. & Kleck, 2005). Most previous research has examined gaze processing using forward (direct) facing heads presented in the observer's central visual field. However, in many real world situations, for example when interacting within a group, we must judge gaze-direction using only peripheral vision. Indeed, inasmuch as the main function of human peripheral vision is to direct eye movements towards salient stimuli, and that a face looking at us is highly salient, we might expect gaze-direction processing to operate effectively when stimuli are viewed with peripheral vision.

Single cell recording from the superior temporal sulcus of macaque monkeys (STS), indicate that there are specific pools of neurons sensitive to direct, leftwards averted and rightwards averted gaze deviations and head rotations ((Perrett et al., 1985). Complimentary functional magnetic resonance imaging (fMRI) studies (Calder et al., 2007) have uncovered comparable regions in the human STS instantiating mechanisms selective for direction of gaze. Pools of neurons that activated in response to presentation of direct or averted gaze were adapted (i.e. their activity was reduced after prolonged exposure) and were associated with a corresponding shift in behavioural responses. Specifically, the perceived direction of gaze shifted away from the adapted direction (i.e. after leftwards adaptation, leftwards gaze directions appeared more direct). Building on these results, it has been suggested that humans process gaze using a multi-channel system, with at least three separate channels coding direct, leftwards and rightwards gaze deviations (Calder et al., 2008).

Signalling of direct gaze is particularly important, informing us when another person's attention is directed towards us. Gamer and Hecht (2007) report that there is a fairly broad range of gaze directions that an individual perceives as being directed at them; a range referred to as the "cone of direct gaze" (CoDG). Using a categorisation technique, Ewbank et al. (2009) showed this CoDG to be broad (8-9°) and, under conditions of uncertainty, humans have a prior expectation that gaze is directed towards them (Mareschal, Calder, & Clifford, 2013; Mareschal, Otsuka, & Clifford, 2014). The latter study induced uncertainty by adding luminance noise to the eye-region of face stimuli and found that observers' perception of gaze-direction was shifted towards "direct". This effect also occurred for turned heads (i.e. where head orientation and gaze direction were mismatched) presenting further support for a prior for direct gaze, rather than a shift in strategy (e.g. observers simply reporting head orientation when uncertain about gaze direction).

Perception of direct gaze, or the feeling of being "looked at", has been a focus of much research into gaze perception. For example, it has been shown that males who have high levels of social anxiety are more likely to feel they are being looked at (Jun et al., 2013) and participants are better at recognising faces exhibiting direct than averted gaze (Macrae, Hood, Milne, Rowe, & Mason, 2002). Given the social significance of direct gaze and that peripheral vision guides future saccades to salient objects; it would be useful for our peripheral vision to rapidly detect being "looked at" so that possible threat can be detected. Senju and Hasegawa (2005) have also shown that presentation of a face exhibiting direct gaze delayed detection of a peripheral cue, suggesting that this is a stronger attention holding cue. Taken together these studies highlight the importance of the perception of being looked at, though how this might occur in the periphery is unclear.

Gaze direction is not derived exclusively from the eyes but also from the orientation of the head. An early example of this is the Wollaston illusion (Wollaston, 1824), where identical eyes appear to be gazing in different directions when placed in two differently oriented heads. Research into the effect of head rotation on perceived gaze direction has generally been divided into those finding that gaze direction is biased either towards the direction the head is facing (attraction) or away from the head rotation (repulsion). For example, Todorovic (2009) manipulated the eccentricity of facial features from the centre of schematic faces (i.e. shifting the eyes, nose and mouth to one side of the face), while keeping the iris eccentricity constant. It was found that shifts in face eccentricity caused the perceived direction of gaze to shift in the same direction (attraction). This effect has also been found using manipulated photographs of real faces as stimuli (Langton, Honeyman, & Tessler, 2004). In contrast to these studies that used artificial stimuli, (Anstis et al., 1969) found that the perceived direction of gaze of a “looker” demonstrator was repulsed from the direction of the head.

Otsuka et al. (2014, 2015) resolved the above conflicting results by proposing a dual channel system where head rotation can exert both an attractive and repulsive effect on perceived gaze. Under this proposal the repulsive effect arises from the rotation of the eye region and the attractive effect from the global head rotation. This is based on the fact that the studies that reported attraction used stimuli where the same eyes were inserted into rotated heads, whereas those that reported repulsion used naturalistic “turned head” stimuli, where the eye region rotated with the head. In this case, head rotation causes a corresponding rotation in the eye region such that the amounts of iris and visible sclera change, leading to a shift in the perception of gaze direction. Otsuka et al. (2014, 2015) found that when only a small window around the eyes was visible, there was a clear repulsive effect of head rotation but that this effect was weaker in a whole head view condition. From this, the authors proposed a two-channel system, where rotation of the eye region exerts a strong, repulsive influence on gaze and the global head rotation exerts a weaker attractive effect, such that the overall effect is one of repulsion.

Here, we examine how people combine head-orientation and gaze-deviation when judging gaze-direction in their periphery. Peripheral vision differs from foveal vision in two essential ways: decreased spatial resolution and increased crowding. Perception in the periphery is poorer for a variety of tasks that require the recognition of fine detail, such as letter recognition (Chung, Mansfield, & Legge, 1998) and numerals (Näsänen & O'Leary, 1998). For isolated stimuli this reduction in spatial resolution is consistent with reduced cortical magnification (Duncan & Boynton, 2003) (the numbers of cortical neurons representing $1mm^2$ of visual space). A quite independent limit on our peripheral vision is set by *crowding*: our inability to recognize objects, such as letters, when they are presented surrounded by "clutter". Under crowding, features of objects and clutter can be erroneously bound together resulting in object mis-identification (Dakin, Cass, Greenwood, & Bex, 2010; Mareschal, Morgan, & Solomon, 2010; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Despite its limitations, peripheral vision allows us to effectively plan saccades by signalling the location of salient stimuli, allowing attention to then be appropriately deployed at fixation (Itti & Koch, 2000).

Most research into the processing of peripherally presented faces has focussed on observers' perception of facial emotion. Emotional information attracts attention when it is presented in the periphery (Calvo & Lang, 2005), suggesting that processing of emotion is preserved even under conditions of degraded visual acuity. Consistent with this, it has been shown that participants are quicker and more accurate at discerning the emotion of a face than its gender, when presented in the periphery (Bayle et al., 2011). This is particularly relevant as it has been shown that whether a face's gaze is directed towards, or averted from, the perceiver modulates the emotion that is perceived (Adams Jr. & Kleck, 2005). There is a suggestion that eyes are more poorly processed in the periphery compared to other elements of the face, particularly the mouth. For example, happy emotions with a distinctive mouth expression are more easily recognized in the periphery than emotions such as fear or surprise, which are conveyed by the eye region (Calvo et al., 2014).

The perception of head and eye rotation in the periphery has been quantified in terms of an individual's ability to resolve head and eye deviations with eccentric fixation. Loomis et al. (2008) tested participants' ability to identify both head rotation and eye deviation, separately, using real face stimuli. When participants indicated the head rotation of a demonstrator using a graspable pointer, performance was near identical between 0° and 45° eccentricity and still showed a linear relationship between actual head rotation and perceived direction at 90°. In contrast, when participants had to indicate on a horizontal scale, the direction of gaze of a photo of a demonstrator's face on a computer screen, their responses tended towards direct above 8° retinal eccentricity, suggesting they were relying on the head direction (which was always direct), rather than accurately reporting the eye deviation. Although this would be expected from a reduction in spatial resolution causing a loss of fine detail around the eye region, the authors suggest there may be an additional role of crowding on peripheral processing of gaze. A recent study reports that direct gaze can be processed in the periphery without requiring attention, whereas averted directions cannot. In their study, Yokoyama et al. (2014) show that participants can discriminate between a direct and an averted gaze but not between leftwards and rightwards averted while their attention is devoted to a central, letter discrimination task. However, this was performed using forward facing heads that may facilitate the processing of direct gaze and diminish that of averted gaze. A similar, more recent study has also shown limitations on peripheral processing of gaze (Palanica & Itier, 2015). The authors report that participants were quicker and more accurate at discriminating direct from averted gaze for faces viewed in the fovea compared to in the periphery. They also report a drop off in discrimination performance past 6° eccentricity. In addition, reaction times were faster when participants viewed forward facing heads with direct gaze in the periphery, suggesting an important role for head rotation in the periphery. Taken together these findings indicate that perceived gaze is not independent of head rotation but exactly how these cues interact in the periphery is unclear.

Here we measured observers' judgement of gaze direction for a range of combinations of head rotations and eye deviations (of the iris and pupil within the sclera), when viewing the face directly (central-view condition) and when the face is presented in the periphery. Given both the reduction in spatial resolution and increase in crowding that will result from peripheral presentation, we expect that the detailed information from the eye region will be lost. This could influence perceived gaze direction by changing the relative weightings of head and eye information; as eye saliency is reduced, the weighting of the eye region in combination with the global head rotation may be reduced, leading to a concomitant reduction in the repulsive bias of the eye region. We also expect that as the information from the eyes decreases, the prior for direct gaze could exert more influence on perceived gaze direction, leading to a greater number of "direct" responses. However, this only holds if the prior for direct gaze (shown for central vision) influences peripheral perception of gaze.

In order to quantify changes in performance with peripherally viewed faces, we applied a psychophysical model to the perception of gaze (Mareschal et al. 2013). The model accounts for performance on the categorization task using three parameters: (a) the bias of perceived direct gaze (the gaze deviation that observers judge to be direct; this value is 0 if there is no bias). (b) The gaze directions at which observers respond equally either direct or leftwards/rightwards; known as the category boundaries. From these values the range of directions over which participants will perceive gaze as direct can calculated. (c) An estimate of the noise associated with the gaze perception process. Given that peripheral perception is limited by both spatial resolution and crowding we would expect an increase in noise as eccentricity increases. An increase in category boundaries as internal noise increases would be predicted by a prior for direct gaze, as gaze would be categorised as direct more often across a wider range of gaze directions under conditions of greater uncertainty.

Methods

Participants

Two authors, JF and IM, and fifteen naïve observers (undergraduates at Queen Mary University of London) participated in this experiment. All participants had normal or corrected to normal vision. Methods were approved by Queen Mary's ethics committee and participants gave written informed consent to take part in the study.

Apparatus

Stimulus presentation and data collection was controlled by a Dell XPS laptop, running MatLab software (MathWorks Ltd) with Psychophysics toolbox installed (Brainard, 1997). Stimuli were presented on a Dell LCD monitor (1440 x 900 pixels, refresh rate 60 Hz). At a viewing distance of 57cm, one pixel subtended approximately 1.8 arcmin.

Stimuli

Four synthetic, greyscale head stimuli with neutral expressions, were generated using Daz software (Daz Productions, figure 1 top row.). The heads were either forward facing or rotated to the left or to the right using FaceGen software (Singular Inversions Inc.). The original eyes were removed from the Facegen 3D models and we inserted greyscale eye stimuli created using Matlab that allowed us to control the horizontal and vertical deviations down to the nearest pixel. A small amount of vergence was added to each eye stimulus, such that the pupils in both eyes converged on a point located 57cm away (viewing distance). Face stimuli subtended on average 9 x 15 degrees of visual angle. Two female faces (one example shown in figure 1) and two male faces were used throughout the experiments.

Procedure

Gaze categorization: Five head rotations were used: forward (facing the participant), and rotated by either 15^{o} or 30^{o} to the left or right of participants. Below we adopt the convention of assigning leftwards (head rotations and gaze deviations) negative values. For each head rotation, nine gaze deviations were tested spanning 20^{o} to the left to 20^{o} to the right, in steps of 5^{o} (i.e. -20^{o}, -15^{o}, -10^{o}, -5^{o}, 0^{o}, 5^{o}, 10^{o}, 15^{o} and 20^{o}). Participants were required to classify the overall direction of gaze as either directed towards them, to their left or to their right. Each trial began with a grey screen presented for 200ms, then the stimulus appeared for 500ms, followed by a grey screen for a minimum of 200ms, after which point the participant responded using the 'j' 'k' and 'l' keys on the computer keyboard to indicate their responses as "leftwards", "direct" and "rightwards" respectively. The next trial began after the participant had given their response. For eccentric fixation conditions a fixation dot was constantly present, level with the centre of the face. No fixation point was presented for the centrally presented faces. Gaze offsets for each trial were determined using a method of constant stimuli. Within a run each head rotation and eye deviation combination (of the 5 x 9 = 45 possible) was presented for each of the four facial identities tested, totalling 180 faces in one run.

Eccentricity: In order to examine the effect of stimulus eccentricity, gaze categorization was measured in a central-viewing condition (observers looked directly at the face, eccentricity = 0 degree) as well as two eccentric-viewing conditions where the participants fixated on a point either (a) 6 degrees of retinal eccentricity from the centre of face (approximately 1.5 degrees to the left or right of the faces' ear) or (b) 9 degrees eccentricity from the centre of face (approximately 4.5 degrees to the left or right of the faces' ear). In the main experiment, the stimuli always appeared in the centre of the screen, with observers fixating to the left or right of the face in the eccentric viewing conditions. Participants completed three runs for each fixation condition, in a random order. Observers (apart from JF who performed all conditions) were randomly assigned to either the leftwards or rightwards eccentric condition, counterbalanced so that we obtained nine sets of data for each eccentric fixation and seventeen for the central viewing condition.

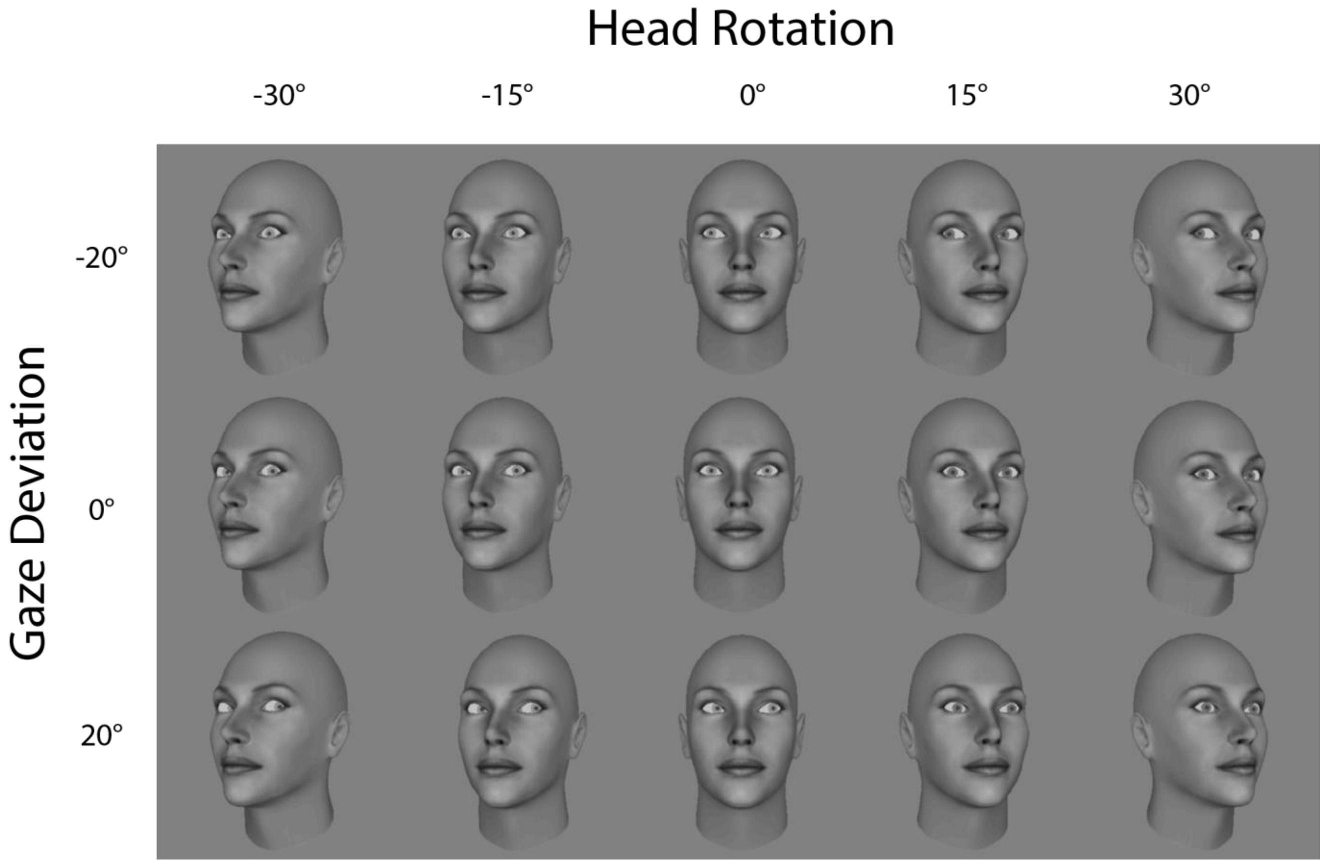

Figure 2:1. Sample female face displaying three head rotations and three gaze deviations

. Faces were viewed centrally (central-view: eccentricity =0 degrees), and peripherally (eccentricity=±6 degrees and eccentricity =±9 degrees).

Results

Categorization of “direct” responses

Figure 2a plots the proportion of responses falling into the three response-classes, averaged across all participants and plotted as a function of gaze-deviation. Observers’ responses to the gaze deviations are as follows: their “leftwards” responses are plotted in blue, their “direct” responses are in black and their “rightwards” responses are in red. Panels are arranged by varying fixation eccentricity (across rows) and head rotation (across columns). Averaged “leftwards” and “rightwards” data were fitted with logistic functions, and direct responses with a simple combination of these functions (1 minus the sum of the “leftwards” and “rightwards” functions; e.g. Ewbank et al. (2009), Mareschal et al. 2013).

There are two main effects to note from these data: (1) when a forward facing head is viewed in the periphery, observers make “direct” responses over a wider range of gaze deviations (black curves in middle column of figure 1) and (2) when a rotated head is viewed in the periphery, observers decrease their “direct” responses (grey highlighted plots) but still respond “leftwards” and “rightwards” to the left and right gaze deviations, suggesting that they are not simply reporting the head rotation.

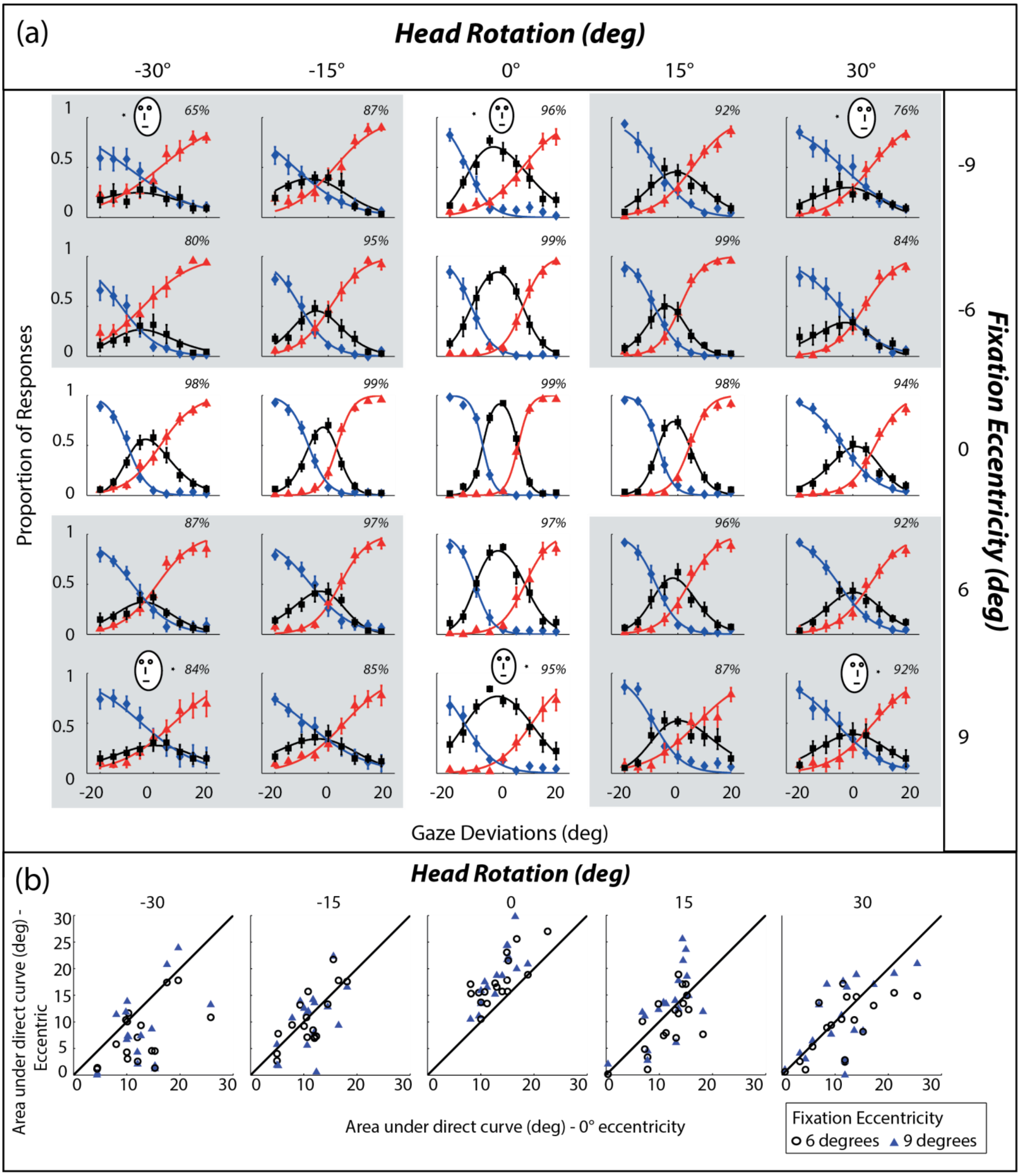

Figure 2:2 Summary of gaze categoristation results.

- (a) The proportion of "leftwards" (diamonds), "direct" (squares) and "rightwards" (triangles) responses, averaged across all participants, plotted as a function of the gaze deviation tested. Error bars represent +/- 1 S.E.M. Each column shows all data for one head rotation and each row plots all data for

one fixation condition (negative values = leftward). Panels shaded grey show data collected with peripherally-viewed turned heads. Schematic insets illustrate head rotation /observer fixation combinations for the corresponding panels. Percentages show the variance explained for each model fit. (b) The area under the curve for "direct" responses, for the central-view condition (eccentricity=0 degrees) plotted against both the near (circles) and far (triangles) fixation conditions. The different fixation directions (left or right) are plotted in the same panel as a function of head rotation. The black line is the line of equality; points above this have a greater AUC in the eccentric conditions than with central-presentation.

In order to quantify the changes in "direct" responses as a function of head rotation and eccentricity, we calculated the area under the curve of direct responses (e.g. area under the black curves in figure 2a). This gives us a measure of how often the participant perceived gaze to be directed towards them, across all gaze deviations. Figure 2b shows, for each participant, the area under the curve (AUC) for their central-view condition (x-axis) plotted against the AUC for both the near (black circles) and far (blue triangles) eccentricities, for each head rotation. Data have been combined into two conditions, 6 and 9 degrees from fixation, independent of fixation side. Points above the equality line indicate that observers responded "direct" more often when the stimulus was in the periphery and data below the equality line indicate they responded "direct" less often for stimuli in their periphery.

A two way, 5x3, within subjects ANOVA was conducted to look at the effect of head rotation and retinal eccentricity on AUC for direct responses. For the purpose of this analysis (and all ANOVAs in this paper) the data from the four peripheral fixations (± 9 degrees and ± 6 degrees) were combined to create two conditions: one for 6 degrees and one for 9 degrees eccentricity, independent of fixation direction. Since there were no clear differences due to direction of fixation (t-tests comparing both the mean AUC for 6 and -6 ($t(16)=-1.17$ $p=.259$)

and 9 and -9 ($t(16)=-.36$ $p=.723$) degree eccentricities were not significant) this allowed us to maintain equal group sizes across eccentricity conditions. In order to combine conditions, data were "leftwards normalised" such that a leftwards rotated head with a leftwards fixation (congruent) was combined with a rightwards rotated head with rightwards fixation (congruent). The rightwards data were flipped, e.g. a "rightwards" response to a rightwards gaze deviation of +20 degrees became a "leftwards" response to a leftwards gaze deviation of -20 degrees, maintaining the relationship between fixation direction and head rotation.

A significant main effect of eccentricity was found ($F(2,34)=3.52$ $p=.041$ $\eta_p^2 = .171$). Post-hoc Bonferroni corrected comparisons revealed that the area under the direct curve was greater for the 9 degrees eccentricity than the 6 degrees eccentricity condition ($t(89)=-3.59$ $p=.001$)), and that the other two conditions were not significantly different from each other. The assumption of sphericity was violated for both the main effect of head rotation and the interaction so a Greenhouse-Geisser correction was applied to the degrees of freedom for these two tests. A significant main effect of head rotation was also found ($F(2.17,36.84)=24.65$ $p<0.001$ $\eta_p^2=.592$). Post-hoc Bonferroni corrected comparisons revealed that a 0° (forward) head had a significantly greater AUC than all other head rotations (0° > -30° $t(53)=-6.64$ $p<.001$, 0° > -15° $t(53)=-7.75$ $p<.001$, 0° > 15° $t(53)=7.31$ $p<.001$), 0° > 30° $t(53)=7.34$ $p<.001$)).

A significant interaction was also found ($F(4.29,72.93) = 8.40$ $p<0.001$ $\eta_p^2=.331$). In order to investigate this interaction further, three one-way ANOVAs (for each retinal eccentricity) were conducted on head rotation. For the 0 degree eccentricity (central-view) condition there was no significant effect of head rotation on AUC ($F(4,68) = 1.78$ $p=.144$ $\eta_p^2=.095$). For both the 6 degree ($F(4,68)=34.83$ $p<0.001$ $\eta_p^2=.672$) and 9 degree ($F(2.55,43.37)=17.59$ $p<0.001$ $\eta_p^2=.508$, Greenhouse-Geisser corrected) eccentric conditions a significant main effect of head rotation was found. For 6 degree eccentricity Bonferroni corrected comparisons revealed that the 0° (forward) head rotation had a significantly greater AUC than all other rotations (0° > -30° $t(17)=-6.91$ $p<.001$, 0° > -15° $t(17)=-6.08$ $p<.001$, 0° > 15° $t(17)=8.81$ $p<.001$, 0° > 30° $t(17)=8.83$ $p<.001$) and the 15° head rotation had a significantly greater AUC than both the 30° (($t(17)=-4.23$ $p=.001$)) and -30° ($t(17)=3.78$ $p=.001$) head rotations. For the 9 degree eccentricity post-hoc, Bonferroni corrected comparisons showed that the AUC for a 0° rotated head was significantly greater than for all other head rotations (0° > -30° ($t(17)=-5.21$ $p<.001$), 0° > -15° ($t(17)=-4.56$ $p<.001$), 0° > 15° ($t(17)=4.80$ $p<.001$), 0° > 30° ($t(17)=5.85$ $p<.001$)).

Taken together this analysis reveals that (a) for the 9 degree eccentricity conditions the AUC was greater than for the 6 degree and 0 degree conditions and that (b) the AUC for a 0° (forward) head across all eccentricity conditions was greater than for any other head rotation. The one way ANOVAs for each eccentricity reveal that the cause of these two main effects is that for eccentric fixations, the AUC is significantly greater for forward facing heads, whereas in the 0 degree eccentricity condition the AUC does not change across head rotations.

Analysis of bias

We sought to determine whether observers not only changed their number of direct responses, but also shifted these responses as a function of gaze deviation, we measured changes in their bias (e.g. what they perceive as being “direct”). In order to compare our results with Otsuka et al. (2014) (who examined bias in central vision), we recoded the data following their procedure where a direct response is attributed a value of 0.5, a left response is given a value of 0 and a right response is given a value of 1. This allows us to plot the data as a single psychometric function that contains information about the three response categories. We fit a logistic function to these data and take the bias as the gaze deviation corresponding to 50% "rightwards" responses (see Otsuka et al. 2014).

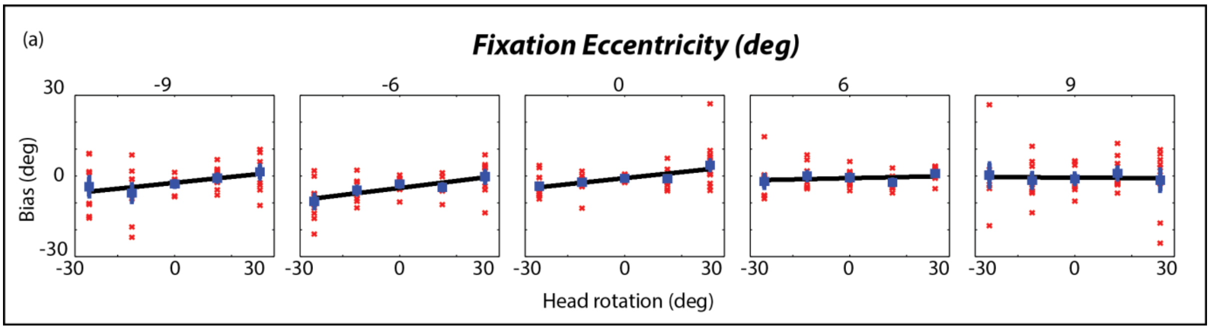

Figure 2:3 - Data show bias in judgements of gaze direction

, averaged across all participants (solid squares), alongside individual data (crosses). Bias is plotted against head rotation for each fixation eccentricity. Error bars represent +/- 1 SEM. The black line is the linear regression to the mean biases.

Figure 3 plots each observer's bias (red points), alongside mean bias across observers (green). In the very few cases (N= 5 out of 425) where the logistic failed to fit observers' data, the data for the condition was excluded from the statistical analysis. A linear regression was fit to the data for each individual's biases across head rotations. Although there appears to be differences in the slopes for the leftwards and rightwards eccentric fixations, no significant differences were found between the mean gradients for the four eccentric conditions (6 degrees v -6 degrees t=1.79 p=.09, 9 degrees v -9 degrees t=1.15 p=.27). Data were therefore combined for the leftwards and rightwards eccentricities giving three eccentricity conditions. The mean of the gradients of these regression lines were compared to a line of slope zero, to determine whether there was a significant effect of head rotation on the bias. We found that for the 0 degree eccentricity condition (direct view), the mean gradient of the regression lines (0.12) was similar to that found by Otsuka et al. (2014) (0.09) and was significantly greater than zero, though the effect size not large ($t(17)=2.16$ $p=0.045$, $d=0.5$, 95% CI=[0.02 0.24]). A positive slope is consistent with a *repulsive* effect of head turn since the bias is in the same direction as the head rotation. For example, in a leftwards turned head, a leftwards gaze deviation is judged as direct (the bias plotted here) which means that the physical gaze is being perceptually repulsed away from the head (see also Otsuka et al. 2014). The mean gradients of the two eccentric conditions did not differ from zero; however there is a (non-significant) trend for this in the periphery, suggesting that the repulsive effect of the eye region is weakened when stimuli are viewed peripherally. These results replicate those of Otsuka et al. (2014) in the fovea, showing a repulsive bias of head rotation on perceived gaze direction. This same effect, however, was not demonstrated in the periphery.

CoDG model

In order to further examine the changes in performance with peripheral viewing, we fitted the model of Mareschal et al. (2013) to each participant's data. The model has three parameters to account for an observer's performance: (a) the peak of direct gaze (the gaze deviation the observer judges most as being direct, e.g. their bias), (b) the width of their category boundaries (between direct and the two averted responses - CBW) and (c) the standard deviation of their sensory representation of gaze (assumed to have a Gaussian distribution). The width of the sensory distribution (SDN) reflects the amount of noise associated with the observers' internal representation of the gaze direction. Figure 4 plots the three parameters, across all conditions for all participants. When fitted to each individual's data, the model accounts for 77.4% of the variance in the data, whereas when fitted to the averaged data it accounts for 90.0% of the variance.

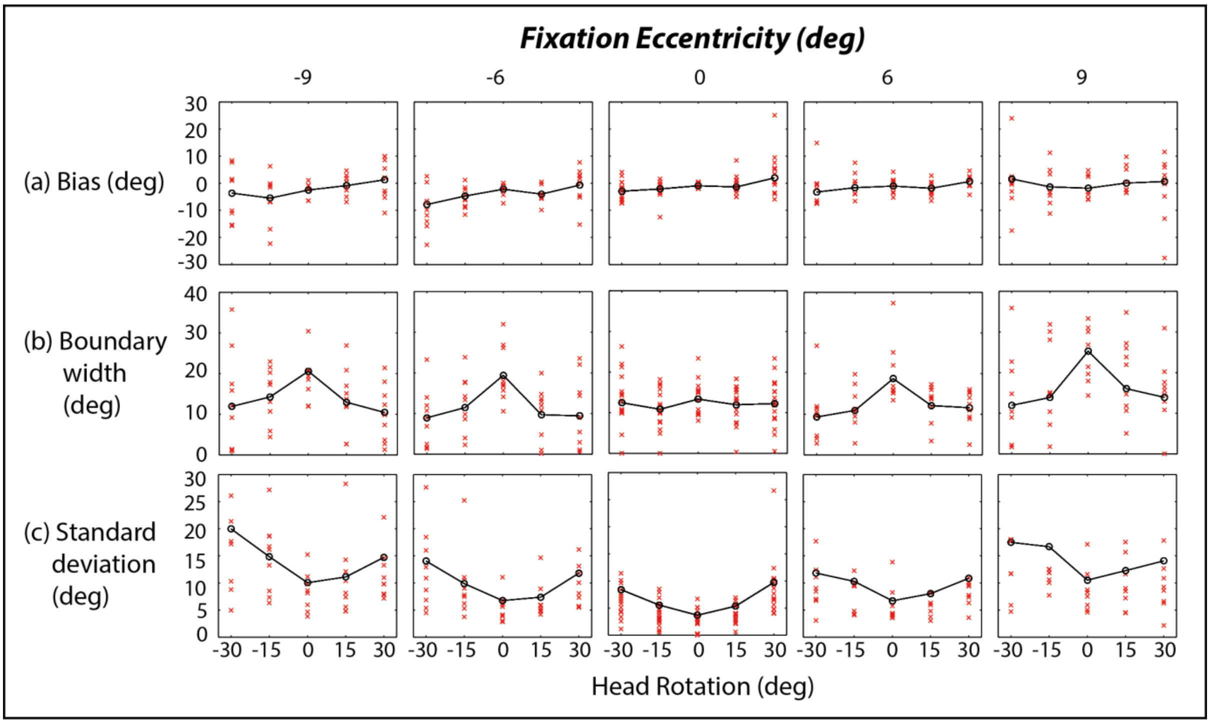

Figure 2:4 - CoDG model parameters

. Each panel plots the parameter values against head rotation for each participant (crosses) and for the averaged data (black lines). (a) Estimates of peak (bias), (b) width (category boundaries) and (c) standard deviation of the sensory representation in the different eccentricity conditions. Each cross is one observer.

Bias results (fig 4a) with the CoDG model are similar to the results obtained from the recoded analysis (fig 3). In order to determine how the effects of head rotation and eccentricity affected the width of the category boundaries (CBW) and the standard deviation of the internal representation of gaze (SDN), data for the far and near eccentricities were compiled as in the AUC analysis, resulting in three eccentricity conditions. Data from participants whose parameter estimates were outliers from the mean estimate (z-scores over 3) were removed for the statistical analysis (4 out of 18).

A two-way, 3x5, within subjects ANOVA was conducted on the CBW data. Significant main effects were found for eccentricity ($F(2,26)=4.873$ $p=.016$ $\eta_p^2=.273$) and head rotation ($F(4,52)=10.376$ $p<.001$ $\eta_p^2=.444$). The assumption of sphericity was violated for the interaction analysis and a Greenhouse-Geisser correction was applied. The interaction was also significant ($F(4.17,54.18)=2.653$ $p=.041$ $\eta_p^2=.169$). When a Bonferroni correction was applied to the post-hoc examination of the main effect of eccentricity, no significant differences between conditions were found. For the CBW data, post-hoc comparisons revealed wider CBW's with a 0° rotated head (forward) than all other head rotations ($p<0.05$), which did not differ from each other.

Three one-way ANOVAs were conducted on the head rotations for each eccentricity condition to look at the interaction between the variables. For the 0 degree eccentricity condition there was no significant difference between head rotation conditions. For the 6 degree eccentricity condition a significant effect of head rotation was found (Greenhouse-Geisser corrected $F(2.12,27.39)$ $p=.008$ $\eta_p^2=.306$). Post-hoc tests revealed that CBW for a 0° (forward) head was significantly greater than the -30°,-15° and 15° rotated heads (0° > -30° $t(14)=-3.54$ $p=.004$, 0° > -15° $t(14)=-5.34$ $p<.001$, 0° > 15° $t(14)=7.80$ $p<.001$); the difference between 0° and 30° was not significant. The one-way ANOVA for 9 degree eccentricity was also significant ($F(4,52)=6.06$ $p<.001$ $\eta_p^2=.318$), the CBW for a 0° head rotation was significantly greater than CBW for -15°,15° and 30° head rotations but not different to -30° (0>30 $t(14)=5.65$ $p<.001$, 0>15 $t(14)=4.7$ $p<.001$, 0>-15 $t(14)=-3.68$ $p=.003$).

The same analysis was also conducted on the SDN data. All comparisons violated the assumption of sphericity so a Greenhouse-Geisser correction was applied. Significant main effects were found for both eccentricity ($F(1.179,15.321)=38.21$ $p<.001$ $\eta_p^2=.746$) and head rotation ($F(2.152,27.975)=10.23$ p<.001 $\eta_p^2=.440$); the interaction was not significant

($F(2.28,29.68)=2.44$ $p=0.1$ $\eta_p^2=.158$). Bonferroni corrected post-hoc tests showed that the SDN for the 6 degree eccentricity condition was significantly greater than that for the 0 degree ($t(69)=-6.80$ $p<.001$) and that 9 degree eccentricity had a significantly larger SDN than the 6 degree condition ($t(69)=-7.79$ $p<.001$). Post-hoc analysis of the head rotation data revealed that the 0° (forward) head was associated with significantly less noise than all other head rotation conditions (-30° > 0° $t(41)=5.18$ $p<.001$, -15° > 0° $t(41)=5.60$ $p<.001$, 15° > 0° $t(41)=-3.80$ $p<.001$, 30° > 0° $t(41)=-6.85$ $p<.001$). As well as this, the 30° head rotation had a significantly greater noise estimate than the 15° head rotation ($t(41)=-4.56$ $p<0.001$). No other significant differences were observed.

Overall there is an increase in CBW in forward facing heads and in eccentric conditions. For all eccentric fixations, a forward facing head causes an increase in the width of the category boundaries, whereas with rotated heads the width of the category boundaries is similar to that in the 0 degree eccentricity condition (where the CBW are not affected by head rotation). This means that a forward facing head in the periphery is perceived as looking at the observer over a wider range of eye deviations than when in the fovea.

There is also an increase in the standard deviation of the internal representation of gaze direction with increasing head rotation and fixation eccentricity, meaning that observers were more uncertain in their judgements under these conditions. Interestingly, these changes are not linked to any change in the cone widths (e.g. compare panels 4b and 4c): observers categorical boundaries for judging whether a gaze is direct or averted (left or right) do not change based on an increase in the uncertainty resulting from head turn and eccentricity.

Spatial Resolution Control

In order to determine whether observers' performance in the furthest eccentric viewing condition was the result of reduced spatial resolution, we M-Scaled our original stimuli so that they were matched in spatial resolution to the 9 degrees eccentric fixation. Nine participants (3 had taken part in the main experiment) performed the categorisation task again for these centrally viewed, M-scaled stimuli. Scaling was done using the formula from Duncan and Boynton (2003): $1/M = 0.065E + 0.054$, where M is the scaling factor and E is eccentricity. The resulting stimulus subtended 3.2 x 5 degrees of visual angle.

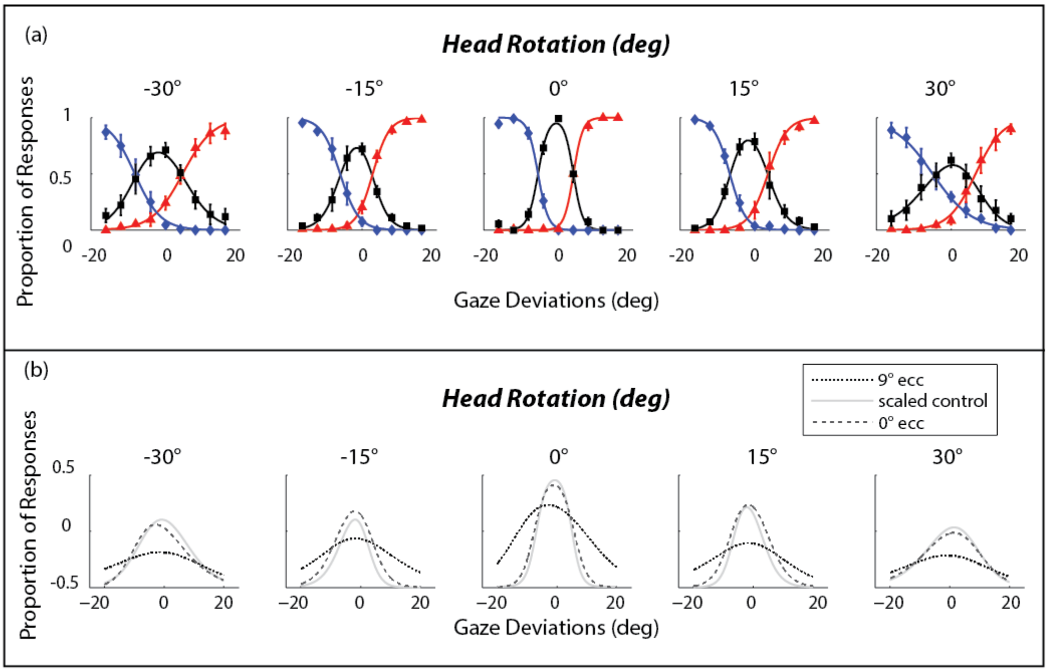

Figure 2:5 - (a) Categorization data averaged across nine observers using M-scaled face stimuli . Each panel shows the proportion of left (diamonds), direct (squares) and right (triangles) responses to each gaze direction for a single head rotation condition. Curves are logistic fits to the data. (b) The "direct" curves for 0 (dashed) and 9 (dotted) degree eccentricity conditions (main experiment) and the M-scaled condition (solid grey).

Figure 5a plots responses as a function of head rotation for centrally viewed M-scaled heads. Figure 5b plots the pattern of direct responses for the scaled control, 0 degree eccentricity and the averaged far eccentric (±9 degrees) conditions. M-scaled data look very similar to the central view data in the main experiment (Fig 5b compare solid and dashed lines). In order to compare the similarity between the M-scaled data and the results from the main experiment, the sum difference between the direct curve fits for the M-scaled faces and the 0 and 9 degree eccentricities (differences in the curves in fig 5b) was calculated for each head rotation. A t-test comparing the mean difference across head rotations revealed that there was a greater average difference between the M-scaled and 9 degree eccentric stimuli than the scaled and 0 degree eccentric stimuli ($t(8)=2.86$ $p=0.02$), suggesting that performance in the periphery is not solely due to changes in spatial resolution.

Discussion

Using a categorization task we find that observers' perception of gaze direction depends both on head rotation and viewing eccentricity. We find that when the stimuli are viewed foveally (direct-view condition), gaze is categorized as "direct" over a broad range of gaze deviations, consistent with earlier reports (Gamer & Hecht, 2007). We also find evidence of a repulsive effect of head rotation that is displayed by the peak of the direct responses occurring at a gaze deviation in the *same* direction as the head rotation. For example, if the peak of direct gaze (i.e. perceived 0°) for a leftwards rotated head is also leftwards (e.g. -3° degrees), this means that the perceived gaze deviation is repulsed away from the head rotation (away from -3° towards 0°), in accordance with the results of Otsuka et al. (2014, 2015).

Using M-scaled foveal stimuli, we have also demonstrated that the changes in peripheral gaze perception are not solely the result of reduced spatial resolution in the periphery. This does not rule out the possibility of other limits on the processing of the gaze direction of peripheral faces, such as crowding. As can be seen from the model estimates of the internal noise on the representation of gaze direction, peripheral faces are associated with more uncertainty than foveal ones.

When stimuli were presented in the periphery, the head rotation largely determined whether the observer classified gaze as direct. When the head was forward facing, the overall number of direct responses increased and the range of eye deviations that were classified as direct also increased. This suggests that the perception of being looked at in the periphery seems to be driven by a head that is forward facing, rather than by any particular cue from the eyes. When heads were rotated, the opposite occurred, with direct responses reducing across all gaze deviations. This result cannot simply be attributed to participants' reporting the direction of head turn, as the 'left' and 'right' responses were not correspondingly affected (e.g. observers never responded only left with a leftwards rotated head and vice versa).

Previous research has suggested that an increase in the uncertainty associated with the processing of a (foveally viewed) face leads to more gaze deviations being perceived as direct (Mareschal, Calder, & Clifford, 2013). Here, we find that the increase in uncertainty due to the face being processed peripherally led to an increase in direct responses for a forward facing head only. When heads were rotated, direct responses were greatly reduced. Although this is not immediately surprising (since the rotated heads never pointed directly at the observer), a few points emerge. (1) Even with gaze deviations that could combine with a rotated head to sum to direct (e.g. -15 degree head rotation with a 15 degree gaze deviation), observers rarely classified this as direct, suggesting that gaze deviation and head rotation don't simply add when presented in an observer's periphery. (2) Given that we report an increase in uncertainty with head rotation in the periphery, this suggests that the prior for direct gaze, shown to exist in central vision with both forward facing and rotated heads, does not hold in the same way in the periphery. It may be that in the periphery other influences (such as, for example, a prior for head rotation) may dominate observers' performance. Given the limits of peripheral vision, it is possible that a prior for "direct" *head* rotation rather than gaze direction (e.g. an increased perception that head rotation is facing the observer), may exist in the periphery. Given the suggestion that forward facing heads attract attention (e.g. Palanica & Itier 2015), a prior for direct head rotation may facilitate the shift in attention to a "direct" head so that the true direction of gaze can be more accurately perceived.

Our results highlight the overriding importance of a forward facing head in the periphery. It has been suggested that two components influence head rotation processing; the symmetry of the outline of the head and the orientation of the nose (Wilson, Wilkinson, Lin, & Castillo, 2000), both of which can be used independently of each other (Langton et al., 2004). Wilson et al. (2000) report that - for centrally viewed stimuli - the average head orientation threshold is low (at around 1.9°), although this increased when discrimination was performed on heads rotated by 30°. For peripherally viewed stimuli, Loomis et al. (2008) found that a high level of sensitivity to head orientation was maintained as far as 90° retinal eccentricity, whereas eye gaze deviation was only accurate to 4° eccentricity (from the closest eye). Our results suggest that observers' may perform some form of a symmetry judgement on the head in the periphery. Given that neurons in the periphery are preferentially tuned to low spatial frequencies (Movshon et al. 1978), these could provide a means for a symmetry judgement, akin to the (large) V4 units proposed by Wilson et al. (2000) in their model of head orientation judgments. Alternatively it has been proposed that the spatial arrangement of internal features allows for direct judgements of facial-symmetry through the use of low spatial frequency horizontal information (Dakin & Watt, 2009).

One intriguing suggestion arising from these results is that of a cascade of information processing, whereby firstly the head outline is assessed as either symmetrical (e.g. forward) or non-symmetrical and then this information influences the width of the category boundaries used to determine whether gaze is direct or averted. For example, if a head is forward facing, it may be that we assume that we are being looked at and therefore don't actively process the gaze. This is consistent with the recent finding that the recognition of direct gaze in the periphery (using forward facing heads) doesn't require attention (Yokoyama et al. 2014). In this case, it may well be that the head cue is processed rapidly and that the observer doesn't make use of the finer information required to process gaze, but simply responds "direct". If so, we predict that response times for categorizing gaze in forward facing heads in the periphery would be faster than when gaze categorization is measured using rotated heads, a finding that has recently been reported by Palanica and Itier (2015).

Our results suggest that discrimination between leftwards and rightwards gaze, particularly in averted heads in the periphery, is still good even out to 9° eccentricity (e.g. fig. 2 bottom left/right panels). This may seem in conflict with reports that gaze discrimination performance falls off between 4° (Loomis et al. 2008) and 6° (Palanica and Itier 2015) eccentricity. However, these differences may simply reflect methodological differences. Loomis et al. (2008) required participants to respond by selecting a number from a range of directions presented in front of them. They report that for stimuli beyond 4° eccentricity, responses were more clustered around direct and did not correspond to the gaze direction presented (reduced accuracy). However, they used forward facing heads for all their stimuli; given our finding that gaze in peripherally viewed forward facing heads is classified as direct over a wide range of gaze deviations, this may explain why most of their responses clustered around direct. More recently, Palanica and Itier (2015) report an increase in discrimination errors between direct and averted gaze for peripherally viewed faces when head rotation and gaze deviation are incongruent (e.g. frontal heads with averted gaze and deviated heads with direct gaze). This is largely consistent with our results; in forward facing heads with leftwards (rightwards) deviated gaze, our observers respond left (right) less often, and in deviated heads with direct gaze, observers respond direct less often. In both cases, this corresponds to an increase in error rate, consistent with Palanica & Itier (2015). Our results differ in that our participants were still able to discriminate between direct and averted at 9° eccentricity, however this may be because Palanica and Itier (2015) presented stimuli briefly (150ms) and required a speeded response, which could have led observers to use the head direction cue, increasing error rates.

The results for the bias using heads in direct (foveal) view show a repulsive effect of head rotation on gaze perception, such that perceived direction of gaze is shifted away from the head rotation. This is consistent with previous findings that head rotation exerts a repulsive influence on gaze direction, mainly due to configural effects of the eye region (Otsuka et al., 2014, 2015). As noted by Anstis et al. (1969) the most notable change in the eye region is the ratio of sclera on either side of the iris when a head rotates. It is likely that this is the cue used to discern the rotation of the eye region that exerts a repulsive effect on perceived gaze direction. Though some studies have reported an attractive effect of head rotation, these either used forward facing eyes inserted into turned heads (Langton et al., 2004; Todorović, 2009) or were confounded by the lighting conditions (Cline, 1967).We do not find a significant repulsive effect of head rotation in the periphery, though there is a potentially interesting (non-significant) difference between the leftwards and rightwards fixation sides (figure 3). The reduction in the bias is most likely due to the changes in weighting of the cues from the head and the eye region. The attractive cue of head rotation (mainly carried by low spatial frequency information, e.g. Wilson et al. 2000) is likely to more strongly influence judgements in the periphery, whereas the repulsive cue of the eye region (requiring higher spatial frequency) would be weakened since resolution decreases with viewing eccentricity.

One function of peripheral vision is to process information in order to plan future saccades (Henderson, 2003). It appears that direct gaze, known to be a strong attention holding stimulus (Senju & Hasegawa, 2005), may have a different effect in the periphery. Our findings suggest that a forward facing head with averted gaze may be more likely to attract attention than a turned head with a physically forward (direct) gaze. These results have interesting repercussions for certain clinical populations for whom direct gaze has been reported to be aversive (e.g. socially anxious or autistic people (Senju & Johnson, 2009b; Wieser, Pauli, Alpers, & Mühlberger, 2009). It is possible that forward pointing faces, viewed in their peripheral vision, might actually exacerbate their sense of being looked at.

Chapter 3 Spatial limitations in averaging social cues

Abstract

The direction of social attention from groups provides stronger cueing than from an individual. It has previously been shown that both basic visual features such as size or orientation and more complex features such as face emotion and identity can be averaged across multiple elements. Here we used an equivalent noise procedure to compare observers' ability to average social cues with their averaging of a non-social cue. Estimates of observers' *internal noise* (uncertainty associated with processing any individual) and *sample-size* (the effective number of gaze-directions pooled) were derived by fitting equivalent noise functions to discrimination thresholds. We also used reverse correlation analysis to estimate the spatial distribution of samples used by participants. Averaging of head-rotation and cone-rotation was less noisy and more efficient than averaging of gaze direction, though presenting only the eye region of faces at a larger size improved gaze averaging performance. The reverse correlation analysis revealed greater sampling areas for head rotation compared to gaze. We attribute these differences in averaging between gaze and head cues to poorer visual processing of faces in the periphery. The similarity between head and cone averaging are examined within the framework of a general mechanism for averaging of object rotation.

Introduction

Social interactions using gaze

The ability to determine where someone's attention is directed is a critical part of human interaction and communication (Kleinke, 1986; Senju & Johnson, 2009a). Information about gaze direction and head rotation are key non-verbal cues that we rely on to determine another's focus of attention (Emery, 2000). These cues can be differentiated with high precision (Anstis et al., 1969; Gibson & Pick, 1963; Wilson et al., 2000) and are processed with specialised neural mechanisms (Calder et al., 2007; Perrett et al., 1985). Although previous research has largely focussed on how these cues are interpreted when presented in isolation, recently there has been increased interest in how they can be rapidly averaged within a group, sometimes referred to as *ensemble coding* (Sweeny & Whitney, 2014).

The perception of gaze direction is critical to human social interactions. It influences the processing of emotion (Adams Jr. & Kleck, 2005), can cue shifts in attention (Driver et al., 1999; Frischen, Bayliss, & Tipper, 2007) and is reportedly abnormal in individuals with conditions such as autism spectrum disorder (Bayliss & Tipper, 2005; Senju et al., 2005). Interestingly it has been found that gaze direction is a stronger cue to attention when the cueing comes from a group of people rather than an individual. For example, Gallup et al. (2012) found that when a participant walked past a group of live actors, the greater the number of people in a crowd performing the same behaviour (e.g. looking upwards), the more likely an individual was to look at the location cued by the group (Gallup et al., 2012; Milgram, Bickman, & Berkowitz, 1969).

Perceived direction of gaze is derived from information about the eyes (position of iris within the sclera (e.g. Ricciardelli et al., 2000) and the rotation of the head (Anstis et al., 1969; Todorović, 2009), with both cues combined to produce an overall percept of gaze direction (Otsuka et al., 2014). Although precision of judgements of gaze direction is generally high when faces are viewed foveally, performance decreases when faces are in the periphery (Loomis et al., 2008, p. 201; Palanica & Itier, 2015). In particular Florey et al. (2015) found that in the periphery, although observers could still discriminate between leftwards and rightwards gaze directions, perceived direction of gaze was more influenced by head rotation. Since there is a dedicated system of neurons for processing gaze direction and head rotation located in the superior temporal sulcus (Calder et al., 2007; Perrett et al., 1985) it may be that the pooling of multiple sources of these cues (e.g. making a judgement about a crowd) also engages specialised mechanisms. That is, there may be a specific mechanism for pooling gaze direction which pools outputs from these face specific regions. This would be consistent with research (Haberman, Brady & Alvarez, 2015) where the authors compared correlations between participants' mean error for face averaging tasks with more basic tasks such as orientation and colour averaging. They found that participants' error for the two face tasks was correlated although neither task was correlated with the low level tasks. The authors concluded that there was no generic mechanism for averaging visual information (e.g. colour, orientation, faces, etc). However, given the diversity of their averaging tasks, a generic mechanism may still exist which averages *directional* cues, regardless of whether they arise from social (faces) or non-social sources. Here we examine whether observers' ability to average the direction of cues of different modalities (rotated heads and rotated cone stimuli) is any different to averaging those of the same modality. If a generic rotation pooling mechanism exists then performance should be the same whether all the elements are from the same modality or not.

Judgements about groups: Equivalent Noise and Reverse Correlation

How do we rapidly extract information about the world? One suggestion is that the visual system extracts *summary statistics* (e.g. mean or variance) of the visual information around us. For example, when perceiving the leaves of a tree, we represent information about the average size, shape and colour rather than information about every leaf (Alvarez, 2011; Dakin, 2015). This has been shown for basic visual properties such as size (Ariely, 2001) or orientation (Parkes et al., 2001). For example, Parkes et al (2001) presented groups of oriented Gabor patterns to participants in their periphery and asked them to make judgments about the orientation of these elements. They found that participants could not accurately report the orientation of any individual Gabor, despite this; they were able to report the average orientation of the patterns in the array above chance. They proposed a form of “compulsory averaging” of visual information in the periphery and suggested that, in peripherally viewed cluttered displays, people cannot extract information about the individual elements.

In order to measure what limits observers’ averaging of information, equivalent noise methods have been developed. Barlow (1957) proposed that the effect of neural noise on detection performance could be treated as being the same as light (going so far as to call such noise, dark light). One could then quantify the amount of light-noise one had to add to the stimulus to match the effect of dark noise. This amount of light is equivalent noise and has since been used to great effect to psychophysically quantify the limits of vision. Notably, Pelli & Blakemore (1990) used it to show that noise associated with photon-absorbance, and not neural noise, limits detection performance under many circumstances, indicating that the visual system is limited by its “front-end” and not later neural processing. Dakin (2001) showed that the paradigm used for quantifying EN could be adapted to study limits on local and global visual processing. He showed by quantifying limits on observers’ ability to make judgements of averages (now called ensemble processing) as a function of external noise (signal variability) one could use the same approach to derive equivalent noise and efficiency in the setting of texture perception.

When applied to averaging (e.g. orientation, motion and size: Dakin, 2001; Dakin et al., 2005; Solomon, Morgan, & Chubb, 2011) this technique assumes that there are two limits on an observer’s ability to extract average information about multiple samples; their *internal noise* and their *effective sample-size*. The internal noise refers to the uncertainty associated with processing any individual element in an array (e.g. how accurate we are at judging the gaze direction of a single face). Sample size refers to the number of elements from the array that the observer’s seem to be combining when estimating the average (e.g. how many faces from a crowd are used to make a judgement). This measure of effective sample size tells us how many samples an ideal observer would have to be using in order to obtain the observed thresholds given constant internal noise. This does not necessarily mean that on *every trial* a participant with a sample size of n is using n elements in their mean estimate, but rather gives us a measure of how efficiently they are combining information across the array. By considering these two separate limits on averaging performance, we are able to show when averaging performance improves, and whether this results from reduced internal noise or an increased effective sample size.

Determining these limits on performance cannot be achieved by looking at changes in sensitivity alone. These two measures of performance can be estimated by measuring changes in observers’ discrimination thresholds (in this case the smallest shifts in mean head rotation that observers can discriminate between left and right fig. 1e) as a function of increasing external noise (in this case changes in the *variability* of head rotation). When external noise is low (i.e. all faces in an array are facing roughly the same direction fig. 1a,b) discrimination is generally good and limited predominantly by the observer's internal noise. When external noise is high (faces are all looking in a wider-range of directions fig. 1c,d), discrimination performance is poor and predominantly limited by how many faces are combined to estimate an average (since the internal noise on any one estimate is now swamped by the external noise). The more faces that are pooled the better the performance. Estimates of an individuals' sample size (n_{samp}) and internal noise (σ^2_{int}) can be extracted by plotting their discrimination threshold (σ^2_{obs}) as a function of the standard deviation of the directions present in the stimuli (σ^2_{ext}) and fitting an equivalent noise function (fig. 1f) to the data.

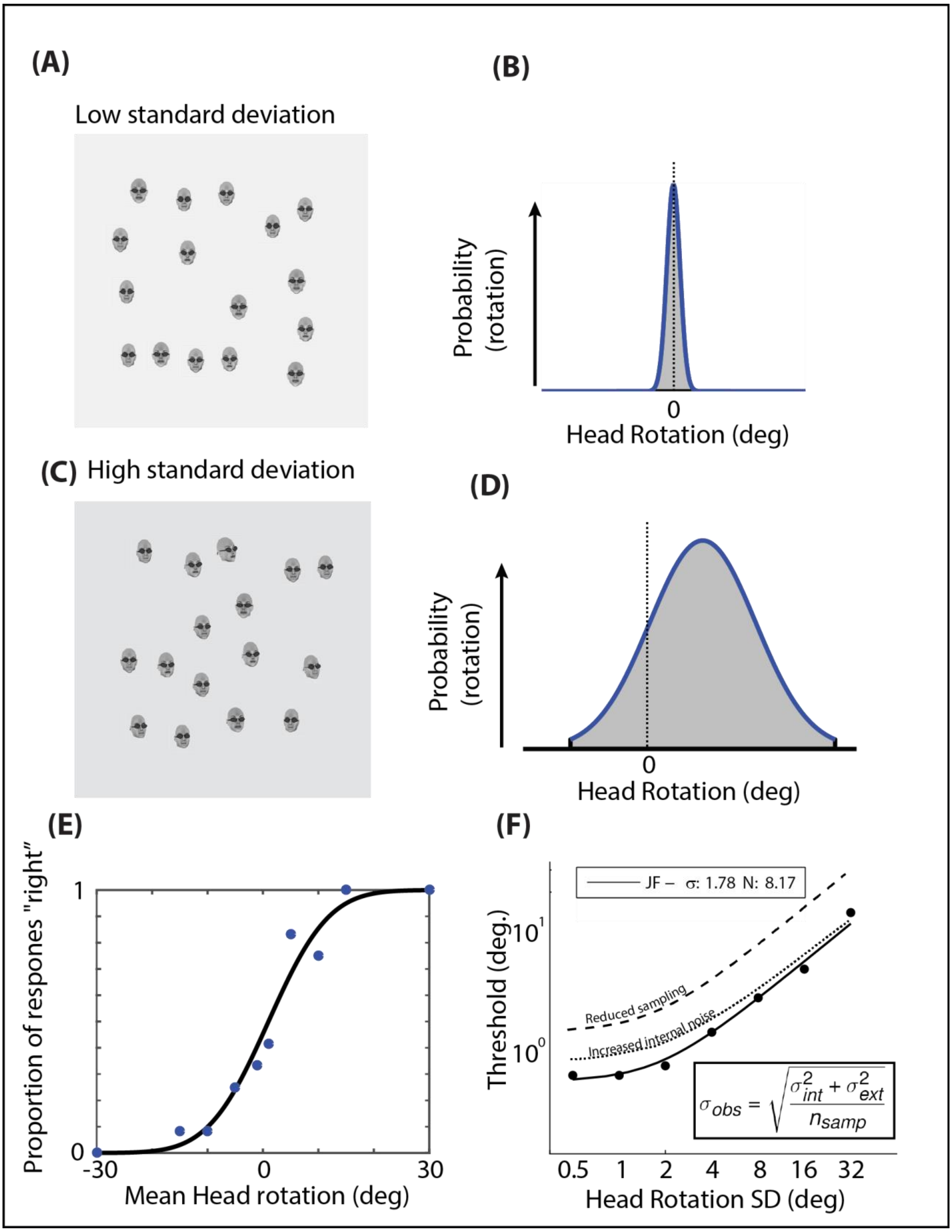

Figure 3:1 Equivalent noise methods.

(A-D): Examples of stimuli and the distributions they are drawn from. (A,B) Low external noise and a mean head rotation close to direct. (C,D) High external noise and a rightwards mean head rotation. (E) - An example of performance on a head rotation task for one observer (JF). Circles show the proportion of times participants responded “right” for each mean head rotation presented. The solid black line shows the best fitting cumulative Gaussian function fit to these points based on maximum likelihood fitting.

Discrimination threshold is taken as the standard deviation of the Gaussian function for this data set. (F) An equivalent noise function for one participant. Reduced effective sample size is characterised by an upward shift in the function, whereas increased internal noise is characterised by an increase in thresholds at lower levels of external noise, but approaching the same asymptote at high levels of external noise. Face stimuli were created using FaceGen Modeller 3.5 (facegen.com) and Poser 10 (my.smithmicro.com/poser-3d-animation-software.html).

Using this method it has been shown that participants use only a subset of samples in an array of Gabor patches to estimate the average (Dakin, 2001, p. 2). Estimates of *subsampling* vary across the literature and depend on the total number of samples in the array, the stimulus property being averaged as well as stimulus presentation duration (Gorea et al., 2014). Estimates range from as many as 80 samples in densely populated arrays of 1024 oriented Gabors (Dakin, 2001) to as low as just 1 out of 100 for motion direction integration in young children (Manning, Dakin, Tibber, & Pellicano, 2014). Using a different method, Sweeny and Whitney (2014) found that the variance in observers' perceived mean gaze deviation of a group of faces decreased when more faces were added to the group (up to four faces). Although their method demonstrates that participants must be averaging to some extent, it does not indicate how many samples (if the population size had been larger) are used. Here we will use the equivalent noise paradigm to examine how efficiently gaze direction and head rotation can be averaged in crowds, a critical first step to understanding how good we are at making (social) judgments about a group.

It is important to note that this procedure is agnostic as to the underlying mechanism. Equivalent noise allows us to characterise observers' performance in terms of e.g. how many samples they are *effectively* averaging. It does not tell us what mechanism allows them to produce this behaviour. Another shortcoming, of equivalent noise methods is that they do not inform us as to *which* samples are pooled in the array when an observer judges the average.

For example we might expect that observers rely more heavily on samples falling close to fixation where resolution is high, but this need not be the case. Reverse correlation methods provide a means of examining this and have been used in a variety of psychophysics tasks (Dakin & Bex, 2003; Mareschal & Clifford, 2012; Mareschal et al., 2006; Neri et al., 1999; Ringach & Shapley, 2004). In conventional reverse correlation, noise patterns are added to a stimulus and the observer is required to perform a (discrimination) task. The observer's performance is then correlated with the noise pattern, on a trial-by-trial basis to determine which noise patterns improved, and which hindered, performance. By adding all the noise patterns that improved (i.e. were positively correlated with) performance and those that hindered (i.e. were negatively correlated with) performance, it is possible to create a map that indicates which parts of the stimulus were used to perform the task.

Averaging of social (faces) stimuli

Although people's ability to average basic visual properties has been well studied, it is useful to expand these methods to broader forms of information. For example, Haberman and Whitney (2009) examined whether participants could average information about facial emotions. They found that when an array of faces with different emotional expressions was briefly presented, participants were able to accurately report the average emotion of the array, suggesting that the ability to infer summary statistics is not limited to basic visual properties. Interestingly, when participants were tested on their memory of the individual faces in the arrays, they only remembered one, suggesting that calculating the average also did not require having conscious representations of each face. The authors suggest that this "ensemble coding" is a rapid process that allows us to quickly extract the gist information from a scene. More recently, Sweeny and Whitney (2014) presented sets of up to four faces around a fixation point and asked participants to report the average gaze direction. Variance in participants' responses decreased when more heads (up to 4) were viewed, suggesting that observers were able to pool gaze directions over space. Importantly, the gaze direction of these faces was produced as a result of the Wollaston illusion (Wollaston, 1824). In this illusion, the gaze direction of the eyes in a face is biased by the rotation of the surrounding facial features; this allows the perceived gaze deviation to be manipulated without changing the properties of the eyes themselves. This suggests that the pooling process required higher level processing beyond simply extracting edges or line orientation (e.g. of the iris in the sclera).

Here we measure the limits of averaging for both gaze direction and head rotation in order to determine whether the mechanism(s) for *averaging* are unique to faces or more generic. Given that gaze deviation is more difficult to resolve than head rotation, particularly for non - foveally viewed faces, we expect that head rotation will be more efficiently averaged (since some heads will necessarily be presented in the observer's periphery). We will also examine whether there is a generic mechanism that pools information, regardless of the object class (using social or non-social). Finally, we will apply a reverse correlation technique to estimate regions of integration: locations in the stimulus array that underlie observers' performance. The reverse correlation maps will also allow direct comparison between sampling efficiencies in the different tasks and stimulus regions of integration (e.g. a smaller highlighted area for conditions with lower effective sample size).

General Methods

The equivalent noise paradigm was used to estimate internal noise and sample-size for pooling gaze direction and head rotation in groups of faces. The data were fit using an equivalent noise function (equation 1) to each individual's discrimination thresholds across a range of (gaze or head) external noise levels.

Equation 1: $\sigma_{obs}^2 = \frac{\sigma_{int}^2 + \sigma_{ext}^2}{n_{samp}}$

Where σ_{obs} is the observer's discrimination threshold, σ_{int}^2 their internal noise, σ_{ext}^2 the added external noise and n_{samp} the effective number of samples used to estimate the mean (e.g Fig 1).

Estimating thresholds*:* The observer's task was to categorise the mean direction of gaze (head or cone rotation) as either to their left or to their right. Thresholds were measured for each level of external noise using a method of constant stimuli (MOCS). Ten mean gaze deviations were presented 12 times each resulting in 120 (randomly ordered) trials per run. The proportion of times the participant responded "rightwards" were plotted against the mean gaze deviation and a cumulative Gaussian function fit to the data using a maximum likelihood estimate method (Fig 1e). Observers completed 3 runs per level of external noise, and thresholds were averaged across the three estimates.

Participants

Participants were two authors, JF and IM and nine naïve undergraduates at Queen Mary University of London. All participants had normal or corrected to normal vision. Not all participants completed all conditions (each participant is allocated a unique set of initials on data figures). All participants gave informed consent and methods were approved by and carried out in accordance with guidelines from the University's ethics board.

Stimuli

Stimuli were sets of sixteen individual faces with each face subtending approximately 2 x 2 degrees of visual angle. On each trial, these were randomly positioned within an 18x18 degree square in the centre of a uniform grey screen with no overlap between stimuli. The identity of each face was chosen at random from a set of four identities, two male and two female made with Facegen (Singular Inversions 2016).

In the crowd stimulus, each face's gaze deviation (head rotation) was drawn from a Gaussian distribution centred on the mean gaze deviation (where negative values indicate leftwards gaze). In most cases, the standard deviations for these distributions were 0.5, 1, 2, 4, 8, 16 and 32 degrees. Due to physical constraints on the gaze stimuli (gaze deviation values that occasionally exceeded the possible range for the human eye), a maximum standard deviation value of 24 degrees was used for all gaze discrimination conditions. The mean rotation values for the MOCS were fixed across observers but changed depending on the stimuli and task difficulty so that a psychometric function could always be fit to the data.

Gaze deviation stimuli

Faces for the gaze deviation conditions were created by first generating greyscale head stimuli with neutral expressions using FaceGen software. The eyes were then removed and the face stimuli were exported to Matlab where new greyscale eye stimuli that allowed for very fine control of gaze deviations were inserted into the face. The pupil and iris were made the same dark grey colour so that the gaze deviation was clearly visible.

There were three different stimulus conditions for the gaze judgement task; one with the whole head present, one with only the eye region visible and one where the eye region was enlarged to approximately match the total area of the heads in the head rotation conditions (fig 2).

For low levels of external noise, gaze deviations were drawn from a Gaussian distribution whose mean was centred on one of the following [-15°, -9°, -6°, -3°, -1°, 1°, 3°, 6°, 9°, 15 °], and standard deviation determined by the noise level (SD= 0.5; 1; 2; 4 and 8). For high levels of external noise, the gaze deviations were drawn from a Gaussian distribution with a mean centred on [-20°, -15°, -10°, -5°, -1°, 1°, 5°, 10°, 15°, 20°] and a standard deviation of either (16, 24).

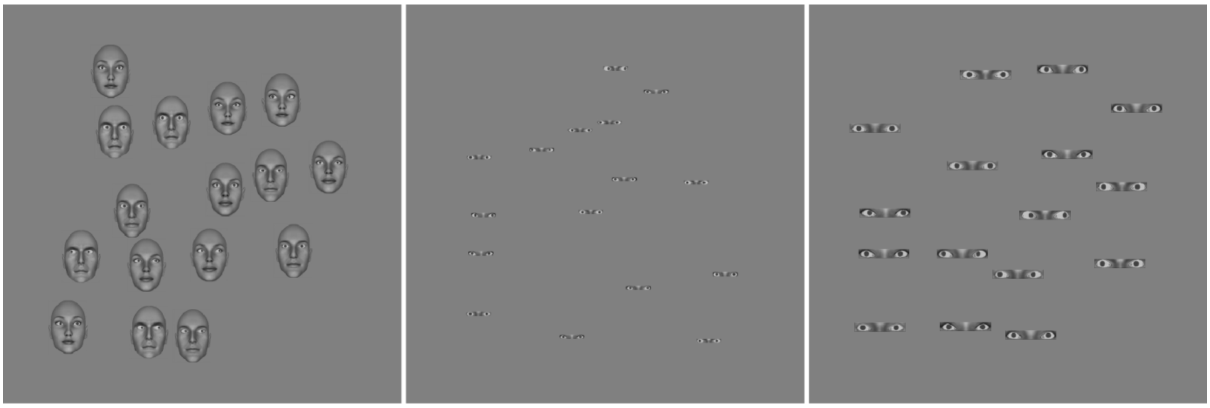

Figure 3:2 - Examples of the arrays of 16 gaze deviation stimuli used

. From left to right: the "full head" condition, the "eyes only" condition and the (enlarged) "eyes only" condition. Here all stimuli have an external noise value of 16°. Face stimuli were created using FaceGen Modeller 3.5 (facegen.com) and MatLab R2015a (uk.mathworks.com/products/matlab).

Head rotation stimuli

To generate the full range of head rotations, the original (forward facing) heads created in Facegen were uploaded to Poser software (Smith Micro 2016), which allowed us to rotate the heads in 3D space along a fixed x-axis. By saving each frame from an animation of the head rotating, stimuli could be exported for head rotations that spanned 180° (leftwards to rightwards) in steps of 0.1°. Black glasses were added to the original Facegen stimuli to remove gaze information.

Observers were required to judge the mean direction of head rotation using the same task. For noise SD values from 0.5° to 4° mean head rotations were [-6°, -3°, -2°, -1°, -0.5°, 0.5°, 1°, 2°, 3°, 6°], when the noise SD was 8° and above, the mean values were [-30°, -15°, -10°, -5°, -1°, 1°, 5°, 10°, 15°, 30°]. The faces were also randomly arranged and placed on a lighter grey background to increase the contrast of the edges of the faces. Data was collected for both upright rotated heads and inverted heads (fig. 3).

Cone stimuli

Non-social directional stimuli were created using rotated cones whose shape was based on a 3D model of a traffic cone. The cone was imported into blender software and its shape edited to remove some of the base and textured to give it a white tip (fig. 3). The full range of rotations was produced in the same way as with the head stimuli. The cone stimuli were approximately the same size as the head stimuli (2x2 degrees of visual angle). In the mixed averaging condition, hybrid stimuli containing both cones and heads were created. The type of cue for each element in the array was randomly assigned as either a head or a cone to avoid participants only responding to one stimulus type.

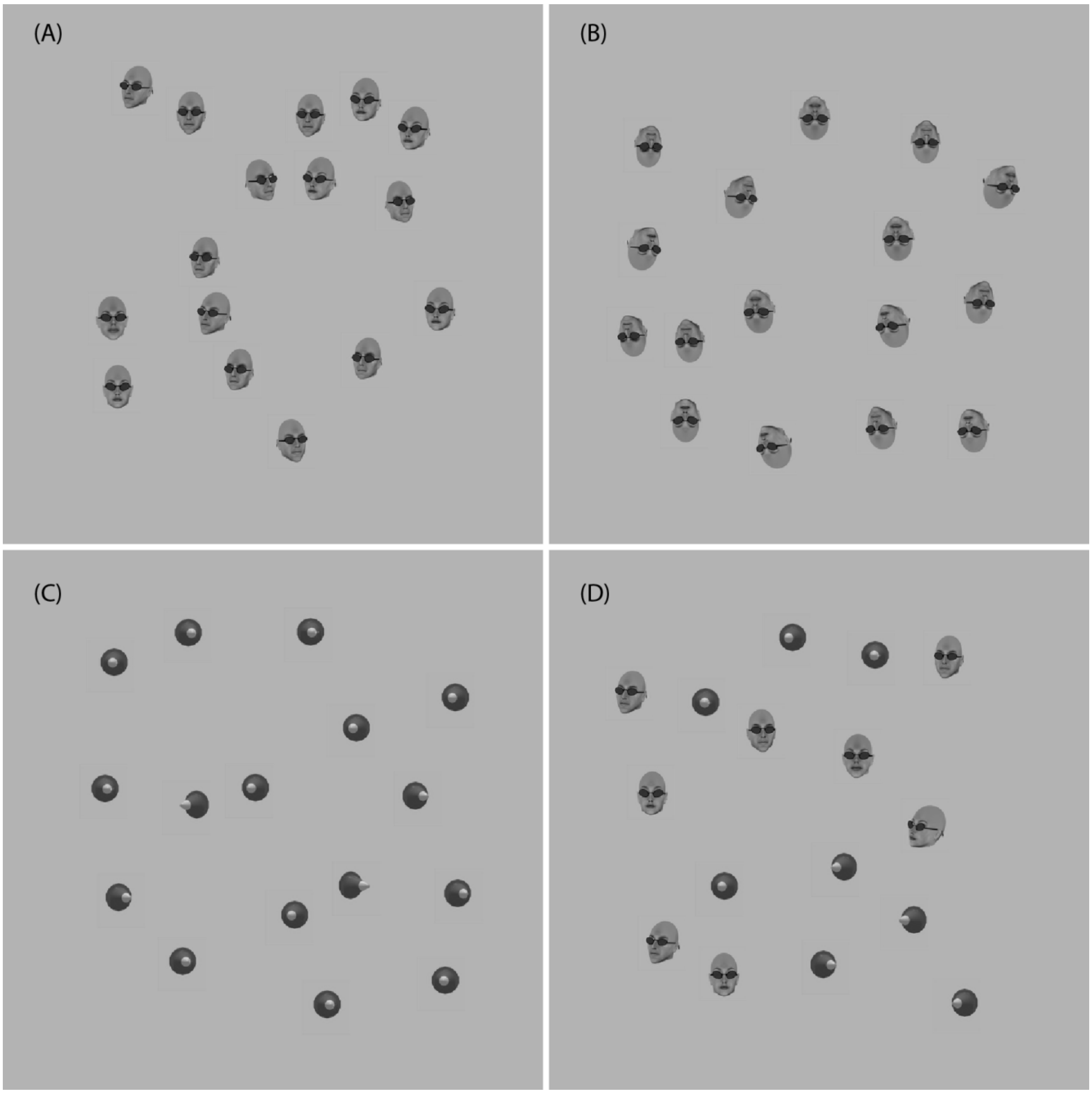

Figure 3:3 - Examples of the four types of stimuli used

to probe judgement of mean object rotation. A-D: Upright rotated heads, inverted heads, 3D cones and a mixture of heads and cones. All arrays have an external noise of 16°. Face stimuli were created using FaceGen Modeller 3.5 (facegen.com) and Poser 10 (my.smithmicro.com/poser-3d-animation-software.html).

Procedure

Each trial began with a blank grey screen (500 ms) immediately followed by a 300ms presentation of the stimulus, followed by a return to the grey screen. Brief presentations were used to ensure participants did not make multiple saccades characteristic of serial averaging.

Participants made untimed responses. The participant's task was to indicate, using a key press, whether the average gaze (head/cone rotation) in the array of stimuli was to their left or to their right. The next trial began as soon as the participant had made their response. No feedback was given.

Reverse correlation

In order to examine which parts of the stimulus contributed to the averaging process we also collected data on the same task in a separate session, optimised for reverse correlation. In the reverse correlation analysis, the most informative trials are when the mean of the stimulus is at threshold as this ensures that participant's response will be driven solely by the subsample of stimuli they use in the image. For example, if the mean of an array is 30° rightwards, most faces will be turned rightwards and a rightwards response from the observer will not be indicative of which faces they based their decision on. However if the mean is 0° then faces are equally likely to be either leftwards or rightwards. In this case the observers' responses will be informative as to which faces they used. In this procedure, the majority of the stimuli (90%) were presented with a mean of 0°. In this case, on any given trial subsets of the faces will be biased to contain either more leftwards or rightwards deviated faces and the observers response will correlate with the subset they used for the task. If an observer has a systematic bias for favouring certain locations within the array (e.g. the centre), then this will be revealed by the presence of a "hotspot" in the reverse correlation map. Catch trials (stimuli with a mean of 10° either leftwards or rightwards) were introduced to ensure that participants were not responding arbitrarily. In all cases, the external noise was fixed at an SD of 16°.

6 participants completed 1200 trials, two authors (JF and IM) and four naïve participants.

Results

Equivalent noise (EN) results

Figure 4 shows the EN function fits to the data for each observer and stimulus type with their corresponding internal noise and effective sample sizes.

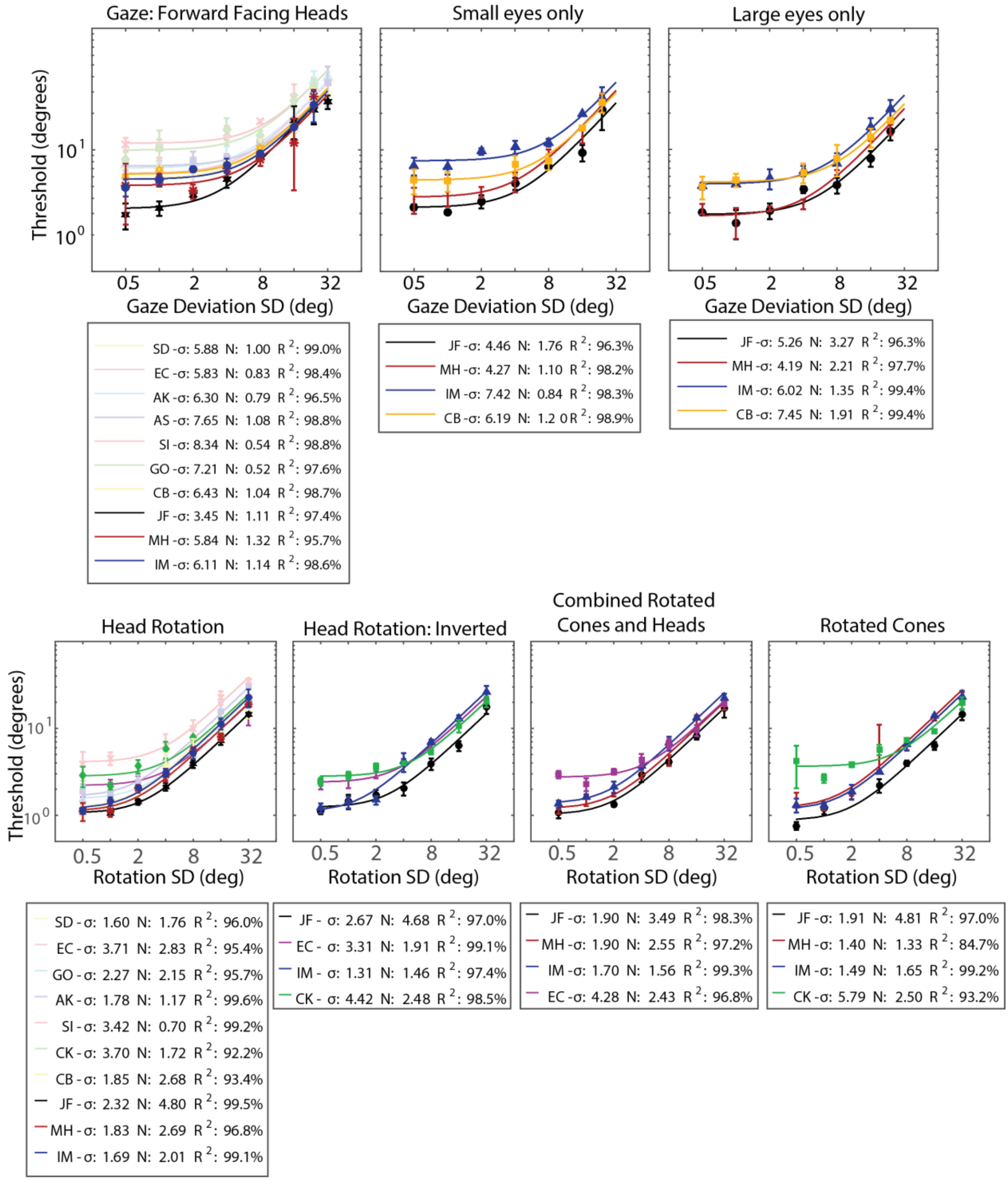

Figure 3:4 - Equivalent noise functions

(lines) fit to the average gaze direction discrimination threshold (symbols, error bars represent +/- 1 SEM) for each participant in all conditions. Top (left to right): the three gaze conditions; full faces, eyes only and large eyes only. Bottom (left to right): upright heads, inverted heads, combined head and cone stimuli and cone only stimuli. R^2 goodness of fit parameters and estimates of internal noise and effective sample size are below each plot. For clarity, data from participants who performed only one of the conditions are semi-transparent.

To compare the two main stimulus conditions (gaze deviation and head rotation in upright faces), independent samples t-tests were carried out on the two equivalent noise parameters. The averaging of head rotation was associated with greater sample sizes (M= 2.31) than gaze direction (M=0.94, $t(9.97) = 3.6$ $p < .001$) and averaging of head rotation was associated with less internal noise (M=2.27°) than gaze direction (M=6.15°, $t(18) = 7.83$ $p < .001$). Taken together these results show that participants were able to average the rotation of heads and cones more efficiently (i.e. using more samples) and with less internal noise than average gaze direction.

Two 1x3 random effects ANOVAs were carried out on the three gaze deviation conditions: full face, eyes only and large eyes only. No significant differences were found between the internal noise estimates for each stimulus type ($F(2,15) = .515$ $p = .61$). A significant main effect was found for the effective sample size of the three conditions ($F(2,15) = 9.99$ p=.002). Post-Hoc Bonferroni corrected t-tests revealed that the “large eyes only” condition was associated with a greater effective sample size (M = 2.2) than the main gaze condition (M = 0.94). Similar 1x4 ANOVAs were carried out on the four rotation conditions: Head rotation, inverted head rotation, cone rotation and the mixed cone and head. No significant difference was found between the four conditions for either internal noise ($F(3,18) = .169$ $p = .92$) or effective sample size ($F(3,18) = .132$ $p = .94$). These results suggest that for both the head rotation and gaze deviation control stimuli, the internal noise associated with processing any item was consistent regardless of the manipulation. Increasing the size of the eye-region and removing the head surround caused observers to sample more efficiently from the array of gaze deviation stimuli. None of the rotation stimuli controls caused any change in observers' effective sample size.

Figure 5 summarises the data for all the stimulus conditions. The average internal noise falls into two groups, with the three gaze deviation conditions associated with higher levels of internal noise than the head/cone rotation conditions. Sample size is approximately 1 for both the main gaze condition and the "eyes only" condition suggesting that there is little to no averaging occurring for these stimuli. In contrast, the head, cone, and mixed conditions as well as the "large eyes" condition all have mean sample sizes greater than 2 suggesting some averaging.

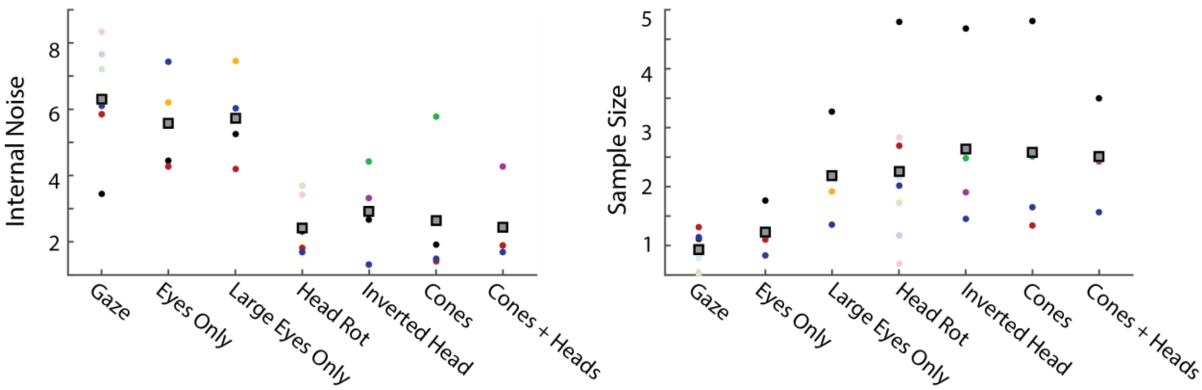

Figure 3:5 –Summary of equivalent noise parameters across all conditions

. Coloured circles show data from each participant (corresponding to the colours in figure 4). Grey squares show the mean across participants.

Mixed stimulus averaging

We found no difference in performance between averaging stimuli of the same type and a mixture of stimulus type (heads and cones). In order to verify that participants were not simply ignoring one type of stimulus (e.g. only using the heads in the mixed condition), we collected additional data on 4 participants for the two highest noise conditions with stimuli made up of both four heads and sets of 2 heads and 2 cones. The total area of the stimulus was reduced in the smaller arrays so that the density of the items was maintained between the first and second experiment. If participants were only averaging across stimulus type we would expect thresholds to increase as the number of samples of the given stimulus type decreased. However, if they were averaging across samples regardless of the stimulus type, then we would expect to see no difference between conditions. Only high external noise conditions that require maximum pooling were tested.

The results from these two conditions are plotted in figure 6. Paired sample t-tests showed no differences in discrimination thresholds between the two stimulus types (16: $t(3) = .434$, $p = .694$; 32: $t(3) = .224$, $p = .837$).

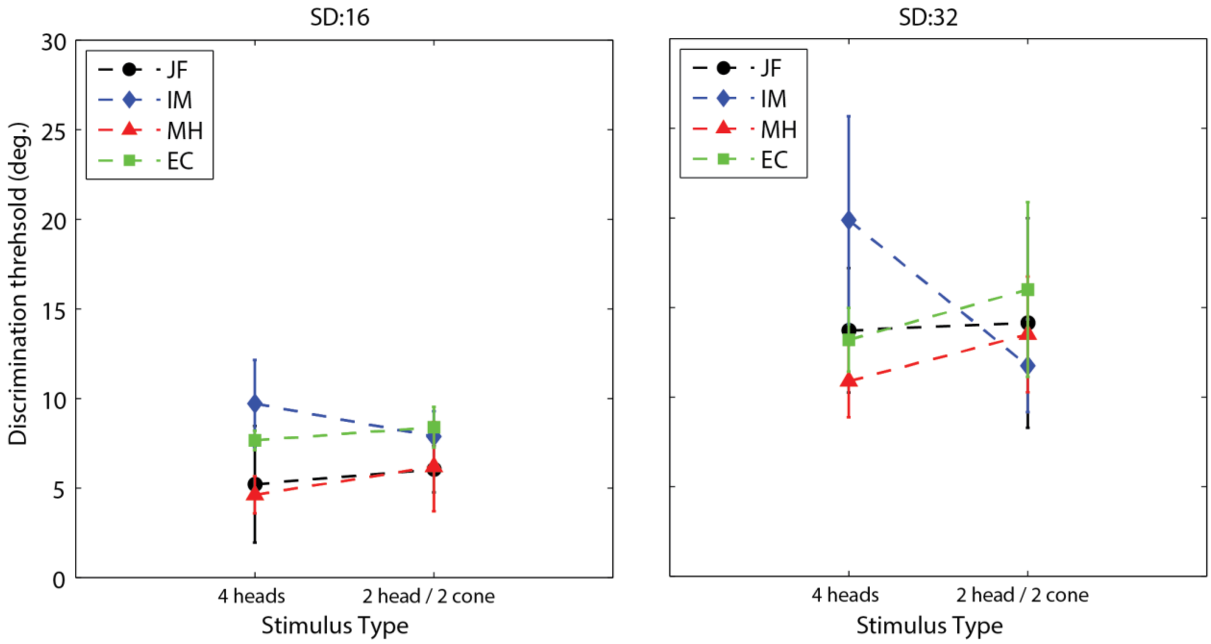

Figure 3:6 –Mean rotation discrimination thresholds for sets of 4 heads and sets of 2 heads and 2 cones . Each line shows data for a separate participant (error bars are +/- one standard deviation) and each graph shows data for a single external noise level.

Reverse Correlation

To generate reverse correlation maps for each participant, the location and rotation of each face in the stimulus was stored along with the participants' response (leftwards or rightwards). Only the trials where the mean direction was 0° were used for the analysis (90% of total).

A reverse correlation map was generated for each participant (fig. 7) by creating an array the size of the total stimulus area for each stimulus trial. All pixels that contained a face within the stimulus trial were assigned a value according to the orientation of the face at that location (1 if rightwards oriented and -1 if leftwards oriented). Pixels in stimulus regions that did not contain a face were not assigned a value. For every trial, we obtained a set of pixel values that were then correlated with the participants' response for the trial. This results in a reverse correlation map where each pixel's value is the correlation coefficient for that pixel across all responses and trials. These images were then smoothed with a Gaussian filter with a standard deviation of three pixels. These images are shown in figure 8 along with the mean of the six participants.

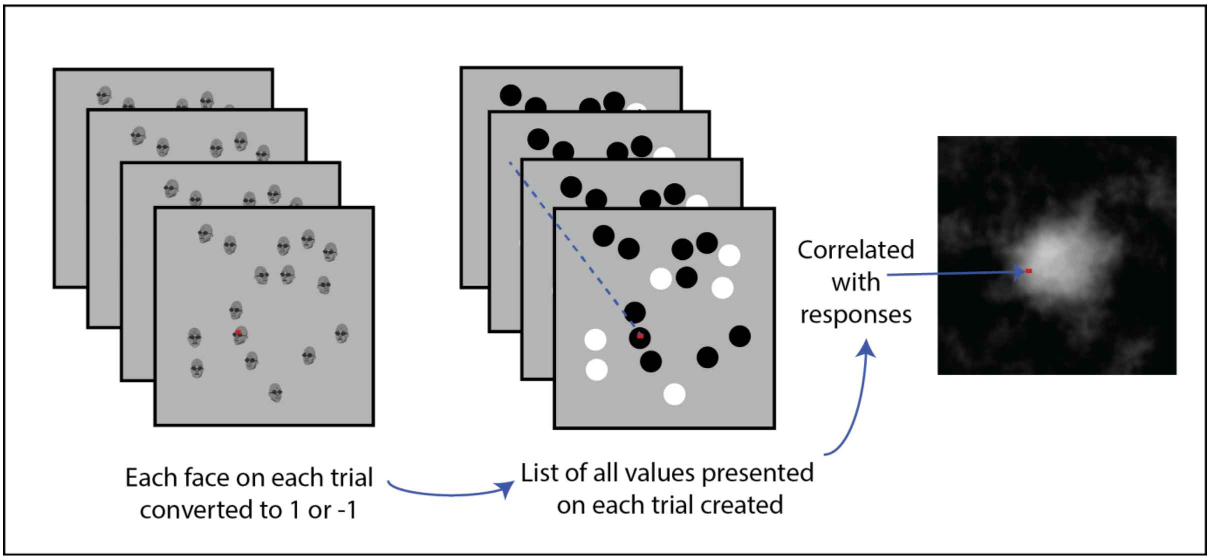

Figure 3:7 Reverse correlation methods.

- On every trial, pixels in the stimulus that were part of a face were assigned a value of [-1 =black] or [1 = white] indicative of the gaze deviation (or head rotation) of the face stimulus at that location (pixels in areas of the stimulus without a face were not assigned a value). At the end of *N* trials, the pixels in the *N* stimuli are correlated with the *N* responses given. Each pixel therefore represents a correlation value, whose magnitude is an index of the influence the faces had on the observer's decision. A reverse correlation map is constructed (rightmost) whereby the influence of a given location in the stimulus is represented by a grey scale.

Two dimensional Gaussian functions were fit to the smoothed maps (fig 8). The width parameter of the 2D Gaussian was greater in the head rotation conditions for both the abscissa ($t(10)=3.73$, $p<0.01$) and ordinate ($t(10)=3.71$, $p<0.01$) directions, suggesting that participants were averaging over a greater area in the head condition relative to the gaze, consistent with the sample size results. There was no significant correlation between the mean width of the two gaussians for each participant and their effective sample size for gaze deviation ($r=-0.35$, $p=0.49$) or head rotation ($r=0.67$, $p=0.15$).

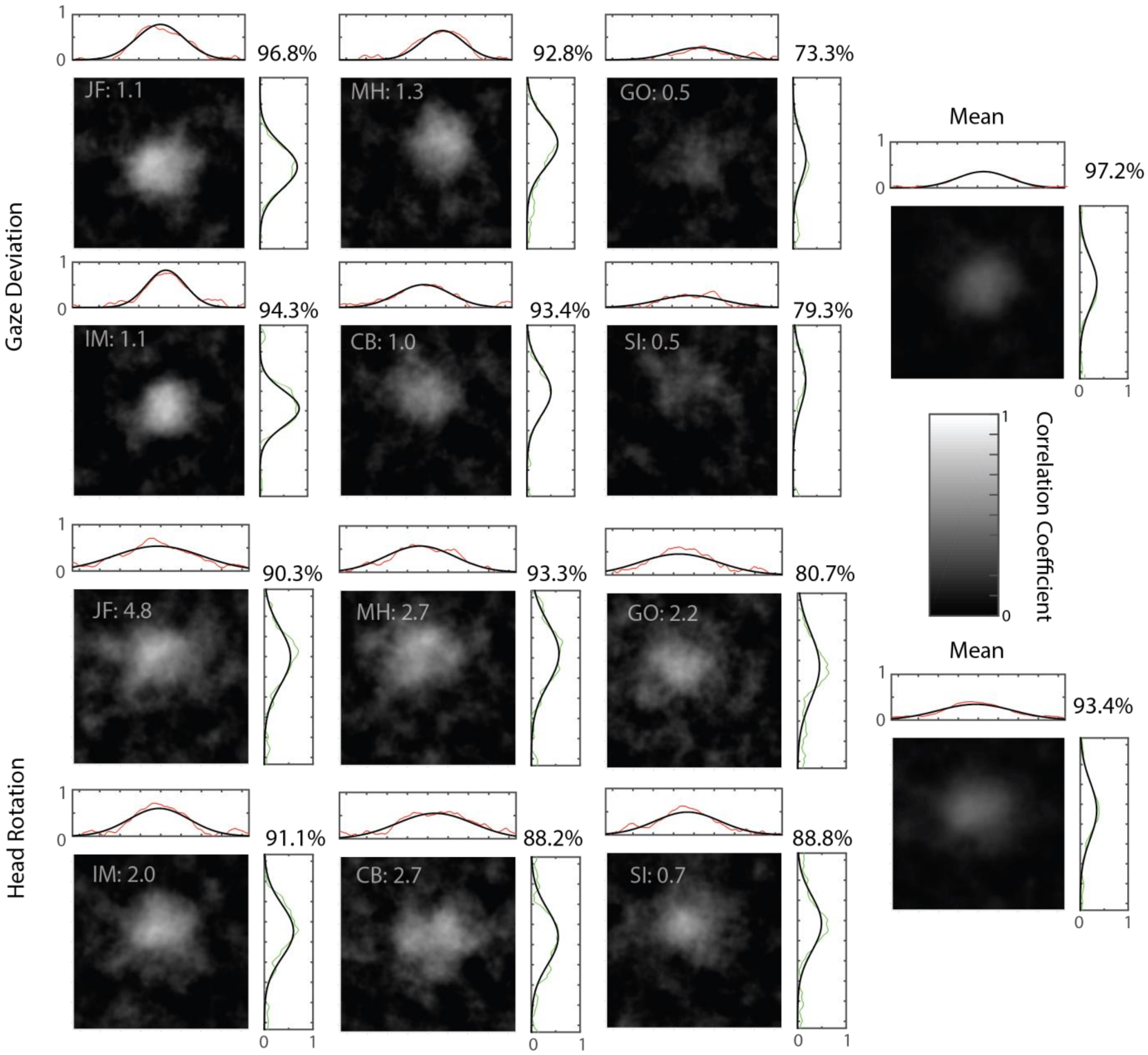

Figure 3:8 - Reverse correlation images for each participant

and the mean across participants for both gaze deviation and head rotation judgements. Grey text in the top left of each plot shows the participant and their sampling efficiency from the corresponding condition in the main experiment. Above and right of each image are the 2D Gaussian fits (black lines) to the smoothed pixel correlation values (red and green points). Percent variance accounted for by fit is shown in the top right of each plot.

General Discussion

We sought to investigate how people make rapid estimates about the average gaze and head rotation in a group of faces. Using an equivalent noise method, we estimated the internal noise and sampling efficiencies associated with pooling these types of cues. In addition, a reverse correlation analysis was applied to determine which elements in the array were being used to perform the task.

The key findings from this study are as follows: **(1)** pooling of gaze deviation is severely limited by size and proximity to fixation since number of samples used in the main experiment is approximately 1, however increasing the size of the eyes (presented alone) improved averaging. **(2)** Pooling of head rotation is more efficient than gaze with observers using approximately 2-4 samples. **(3)** Rotation cues from non-social cone stimuli were averaged as well as heads when matched in size. We also report that observers are able to combine information from head and cone directions. **(4)** Reverse correlation analysis revealed that there was a general trend to preferentially use samples in the centre of an array. In addition, participants used samples from a significantly greater area in the head rotation condition compared to gaze, concomitant with the increase in effective sample size between the conditions. Surprisingly however no clear relationship emerged between the efficiency of sampling of individual participants and the reverse correlation maps, which may result from trial by trial fluctuations in strategies.

The finding of inefficient pooling of gaze deviation is mainly consistent with previous findings. Our results for effective sample size are lower than the majority of studies reporting samples for averaging size (Gorea et al., 2014) and motion (Dakin et al., 2005), however Manning et al. (2014) have observed sampling efficiencies at or below one for motion in children. Given that judging gaze direction of faces in the periphery is much poorer than in the fovea (Florey, Dakin, Clifford, & Mareschal, 2015; Loomis et al., 2008), it is not surprising that averaging gaze direction in a crowd of faces (most of which fall in the periphery) is also poor. The increase in effective sample size in the “large eyes” condition further supports this explanation of reduced peripheral resolution for the gaze condition. Note however that although we report that increasing the eye size brought the effective sample size to a similar level for gaze averaging as for the head averaging condition, this does not reflect averaging performance under naturalistic conditions. Indeed, eyes are always embedded in a (larger) head and therefore heads and eyes will never be matched in size.

Sweeny and Whitney (2014) reported that participants were able to average gaze deviation over multiple faces, though they make no claims about the number of faces being used. This appears inconsistent with our result that participants are not using more than a single sample when averaging gaze deviation. Crucially, their faces did not actually vary in the offset of the pupils; instead they varied the orientation of the features around the eyes, producing shifts in perceived gaze direction as a result of the Wollaston illusion. This may be why their results suggest that up to four faces are being used in some cases, since the important cues (e.g. the orientation of the features around the eyes and overall head rotation), are easier to resolve in the periphery than the eyes themselves and therefore can be used in the averaging process more efficiently.

We find that head rotation was slightly more efficiently averaged than gaze deviation, with the estimated number of samples greater than 2. This approximates earlier reports (Dakin, 2001; Dakin et al., 2005) that suggest that the number of samples a participant uses is equal to the square root of the total number. This improvement in effective sample size is most likely the result of head rotation being more easily discriminated in the periphery (Loomis et al., 2008). Inverting the heads did not have a significant effect on effective sample size nor did averaging arrays of cones, or creating arrays that contained both types of stimuli. Given that it is unlikely that we have a specific mechanism to process cone direction, a more parsimonious account would be the involvement of a generic mechanism that averages the rotation of objects.

It may be surprising that samples from different stimulus categories (heads and cones) are pooled as well as samples from the same category. Since summary statistics are believed to allow the rapid extraction of "gist" from a complex environment it is not surprising that they could be used to extract the average head rotation of a crowd. It is less intuitive for unrelated objects to be pooled together. Chong and Triesman (2005b) have demonstrated that the average of multiple perceptual groups can be extracted at the same time. Participants were able to simultaneously perceive the average size of groups of circles of different colours, though it has been suggested that this process comes with a cost to accuracy (Brand, Oriet, & Tottenham, 2012). It may be that in our case the two types of stimuli are averaged separately and then combined to create a global average. A parsimonious account of our data involves a generic averaging mechanism, however this need not be a high level mechanism dedicated to rotation per se. For example, preliminary evidence suggests that observers are able to average 2D position (Bossi & Dakin, 2014) cues (offsets in spatial position of features that arise as a consequence of 3D rotation). It remains to be seen whether these 2D positional cues are used in our study or whether observers are engaging a specific mechanism for averaging 3D rotation.

The reverse correlation analysis revealed that on average, participants use the central elements in a stimulus to make their decision. Although there was no fixation point it would make sense for participants to fixate centrally as this would be most efficient for processing and is consistent with observers' bias to fixate in the centre of a screen (Foulsham & Underwood, 2008; Tatler, Baddeley, & Gilchrist, 2005; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). Prima facie, this result is inconsistent with findings from Wolfe et al. (Wolfe, Kosovicheva, Leib, Wood, & Whitney, 2015) who found that observers performed equally well on an emotion averaging task, regardless of whether information was available around fixation or not. It is possible that this may reflect differences in the stimuli used; peripheral emotion can be perceived with some degree of accuracy (Calvo et al., 2014) whereas peripheral gaze cannot (Florey et al., 2015; Palanica & Itier, 2015). Alternatively, it may be the case that observers alter their sampling strategy to adapt to the structure of the image. For example, they may be biased towards using foveal information when it is available but distribute their attention further when it is not. Alvarez (2011) used simulations to show that a weighted averaging (with greater weights given to elements sampled with greater precision) would yield more accurate performance than an equally weighted average when samples were estimated for a size averaging task. The individual differences in our reverse correlation data do not suggest that sampling strategy aids the averaging process. For example, both JF and IM have the clearest central bias, with similar sampling areas yet have different sampling efficiencies (e.g. 4.8 vs 2.0 for head rotation). Instead, these maps suggest that the differences in effective sample size between participants are more likely due to other factors such as differences in the way they compute the average (e.g. using a weighted average or taking the maximum of a subset).

The direction of attention of a crowd can be an even stronger cue for attention than that of an individual (Gallup et al., 2012) so it makes sense that there might exist a mechanism for rapidly pooling this information. Our results suggest that it is likely that head rotation is rapidly pooled as this can be done more efficiently than gaze. It is possible that after an initial summary statistic is extracted for head rotation, the individual gaze deviations of the members of the crowd may be scanned in a serial fashion to get a more accurate estimate of the true point of interest.

Chapter 4 Comparing Spatial and Temporal Limits in Averaging Social Cues

Introduction

Observers' ability to extract summary statistics from groups of objects is well established. The perceived mean of low level properties such as orientation (Dakin & Watt, 1997; Dakin, 2001; Solomon, 2010), size (Ariely, 2001; Chong & Treisman, 2005b) and motion Dakin et al., 2005) can be reliably estimated from groups ("ensemble" stimuli). Recently it has been demonstrated that these summary statistics can be estimated over complex, "higher-level" properties, such as facial emotion, identity and gaze direction (Florey, Clifford, Dakin, & Mareschal, 2016; Haberman & Whitney, 2009; Sweeny & Whitney, 2014). Although most research has focused on averaging across spatially distributed arrays of items, observers can also average over temporal sequences (Albrecht et al., 2012; Gorea et al., 2014; J. Haberman et al., 2009; Piazza et al., 2013). We have previously shown that observers' averaging of gaze-deviation over space is limited compared to their ability to average head rotation (Florey et al., 2016). This sets an important limit on our ability to process crowds of faces, as it has been shown that humans are more sensitive to the direction of attention of a group of faces than they are to an individual face (Gallup et al., 2012). The question remains however whether there is a difference in how well people average information in different domains. Specifically, are the limits on averaging stimuli over space the same as those for averaging in time (e.g. when stimuli are presented sequentially), and does this depend on the type of stimulus used?

Perceptual averaging

Although observers are able to estimate average properties from ensembles, they do not behave as though they are using all of the elements available. Dakin et al. (2001) demonstrated that when averaging the orientation of ensembles of Gabor patches, participants performed as if they were using only a sub-set of the total items in the array. Similar results have been found for averaging of other low level properties such as motion (Dakin et al., 2005) and size (Solomon, Morgan, & Chubb, 2011). Recently, Manning et al. (2014) found that children can sometimes perform as if they were using only a single sample to determine the direction of motion of an array of 100 moving dots.

This same sub-sampling effect has also been found for face stimuli. When briefly presented with arrays of either faces with different gaze deviations or 16 heads rotated differently, observers were able to judge the average, though their estimate was based on their (effective) use of a sub-set of items (Florey et al., 2016). Observers were particularly limited when averaging the direction of gaze from a group, in some cases effectively basing their responses on a single face. We created classification images by correlating observers' responses with the distribution of locations and gaze or head offsets presented, which maps the stimulus locations that contributed to observers' judgements. This revealed that participants were biased towards using elements in the centre of the array; an effect that was more pronounced for gaze deviation than head rotation. It has recently been shown that gaze deviation perception is poor in the periphery (Florey et al., 2015; Loomis et al., 2008; Palanica & Itier, 2015), which likely explains the limited contribution of elements falling near the edge of displays.

Spatial limits

In tasks that are less limited by peripheral resolution than gaze perception, spatial integration still suffers from sub-sampling, suggesting there are other limits on spatial integration. One possible limit is the spread of an observers' attention. Chong and Treisman (2005a) found that a dual task that encouraged a spread of attention, improved participants' size-averaging. Although this seems to suggest that a wider spread of attention improves performance on an averaging task, increasing the duration of the stimulus presentation does not. Chong and Treisman (2003) found that reducing exposure duration from 1000ms to 50ms had a little impact on size averaging. Sweeny and Whitney (2014) found that reducing the presentation time of a set of four gaze deviation stimuli from 1000ms to 200ms appeared to *actually* increase the number of elements being integrated. It seems then that there are global limits on integration within spatial ensembles (e.g. distribution of spatial attention and presentation time), that are distinct from limits on the processing of the individual elements within spatially-distributed arrays (e.g. limited peripheral perception).

Temporal limits

Similarly, averaging of visual cues over time is not perfectly efficient. Gorea, Belkoura and Solomon (2014) suggested that when averaging the size of a temporal sequence of circles, participants performed as if they were using up to 4 out of 8 elements. In most sequential averaging tasks, stimuli appear at fixation, reducing the limiting effects of either the perception of any one element in the array (e.g. due to eccentricity), or the effect of spatial distribution of attention. However there are unique factors that influence sequential averaging; notably biases towards favouring particular elements within the sequence. Researchers have used regression analysis to show that observers are biased by primacy (increasing the weighting of items appearing early in the sequence) and recency (increasing the weighting of later items) (Gorea et al., 2014; Hubert-Wallander & Boynton, 2015). Hubert-Wallander and Boynton (2015) examined these effects using different types of stimuli and found stimulus-specific differences in the bias, with face expression and size averaging producing recency effects but position averaging producing primacy effects.

There is some debate as to the reason for these biases. Primacy could result from serial dependencies, a perceptual effect where each element in a sequence is biased to appear more like the item preceding it (Fischer & Whitney, 2014). If each element is influenced by the previous one in the set, responses will be biased towards the early elements, leading to a primacy effect. Alternatively, primacy may result from observers adopting a strategy of ignoring later samples, potentially because they have a limited capacity for integration. An efficient strategy would be to stop adding more information to the average computation, when the resource cost of including it outweighs the potentially improvement in accuracy; such behaviour has been observed in both human and non-human primate observers (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012). One possible explanation for recency is limited attention or memory resources, resulting in early information being ignored or forgotten. Another is adaptive gain control (Cheadle et al., 2014), where elements that are consistent with the expected mean of the sequence thus far are upweighted and those that are inconsistent are downweighted. This type of strategy has been shown to produce recency in both simulated and human data.

Mechanism of averaging

How is averaging performed? The neural mechanism for averaging is not well defined. It is clear that we can perform less than perfect averaging, but beyond that our understanding is limited. Allen, Hess, Mansouri and Dakin (2003)showed that orientation-averaging does not automatically pool estimates from luminance and contrast-defined elements, suggesting distinct averaging mechanisms for both. Haberman Brady and Alvarez (2015) provided evidence against a single generic mechanism for averaging of any type of stimulus. In their study they show that individuals' performance on low level averaging tasks (size/colour/orientation) does not correlate with their performance of averaging high level face stimuli. However, they do report correlations within groups of similar stimulus type (high/low) suggesting some commonality within stimulus type. However evidence for domain-agnostic averaging comes from Florey et al. (2016), who found that observers were equally good at integrating information within a stimulus group (faces) as between stimulus groups (faces and 3D cones). Whether the same mechanisms are employed in spatial and temporal averaging is less well understood. There are necessarily differences in the early visual processing for these two types of stimuli but it seems plausible that the higher level mechanism that integrates multiple elements into a single summary statistic may operate across stimulus types.

Noise paradigms

One method for examining performance on averaging tasks is using an equivalent noise (EQN) procedure that estimates two limits on observers' performance: *internal noise* and *effective sample size*. This is based on the assumption that when observers average, they first estimate the feature-of-interest for individual items in the set (e.g. the orientation of each Gabor in an array) before averaging a sample of these estimates. Internal noise refers to the observers' uncertainty about a single feature. Effective sample size, tells us how many samples an ideal observer would need to average (given the internal noise on each sample) to achieve the observers' level of performance. For a stimulus containing *n* elements, a perfect observer would have no internal noise and an effective sample size of *n,* perfectly averaging the features of all the items into an accurate representation of the mean. In reality, observers display some amount of uncertainty associated with processing of individual items and sub-sample, using only some proportion of the ensemble to estimate the average. To estimate internal noise and effective sampling EQN experiments measure observers' averaging in the presence of different levels of variability of the feature-of-interest. When the variance is low (e.g. Gabors with similar orientation), performance is limited by the internal noise; if they accurately perceive the orientation of one element, they will give a correct response. When variance is high (e.g. Gabors with widely differing orientations), the precision of any one estimate becomes less important - as the variance in the feature will swamp the influence of internal noise on individual elements. In this situation, the number of elements averaged will determine the precision of the observer's response. Modelling performance as variance in the stimulus changes, using an ideal observer, allows one to recover internal noise and effective sample size.

Using the EQN method, previous research into spatial averaging has found that observers effectively use only $\sqrt{n}$ elements for orientation averaging (Dakin, 2001) or even fewer in the case of children averaging motion direction (Manning et al., 2014) or adults averaging gaze and head direction (Florey et al., 2016). Similarly, Gorea et al. (2014) found that observers only sample a sub-set of elements from a set of circles in a sequential size averaging task.

Solomon and colleagues have employed a model related to equivalent noise for integration of orientation and size stimuli (Gorea et al., 2014; Solomon, 2010; Solomon et al., 2011). The key difference in their "Noisy, inefficient but otherwise ideal observer" model of cue integration is that they separate the internal noise term into two separate sources of noise; one that acts before the entire summary is integrated, either on individual stimuli or on "local pools” of subsets of stimuli (early noise), and one that acts at the level of the ensemble code, before a decision is made.

The current study

We have previously suggested that head rotation may be a useful cue to acquire a gist percept of a group of faces (a crowd) and that a serial average of the gaze deviation of individuals may provide a more precise, albeit slower average. Here we measure observers’ ability to average head rotation and gaze deviation both in temporal sequences and across spatial arrays, using EQN analysis. The presentation duration and size of the spatial and temporal arrays will be matched to allow a comparison with an equal amount of processing time available for each. This means that observers can make multiple saccades in the spatial condition, creating more naturalistic viewing conditions for crowd perception (Florey et al., 2016). Under both of these spatial and temporal conditions, we would expect that gaze deviation averaging should be similar to head rotation averaging, as the observer will be able to fixate the faces separately, eliminating the limits on peripheral processing of gaze deviation. Alternatively, observers may not employ an efficient saccade pattern when presented with spatial arrays of gaze stimuli, (i.e. they do not saccade to a new face each time or saccade between faces without processing each foveally) and as a result, they may not improve relative to brief presentations (e.g. 300ms used by Florey et al., 2016).

Below we compare performance across two types of averaging strategies (in space and in time) to determine if individuals who are efficient averagers in one domain are also efficient in the other. This would suggest a generic limit set by the *individual* rather than the *stimulus*. If performance correlates both domains, this would suggest that averaging is limited by the perceptual or cognitive limits of the individual. For example, certain individuals may have high neural noise for processing individual elements, which would limit their averaging of both spatial and sequential ensembles. Alternatively, there may be a shared limit on the number of samples that can be integrated, due to a shared mechanism for integration that occurs beyond the level of visual processing. If there is no correlation, then it is likely that limits on averaging are determined by the manner in which stimuli are presented, either as a result of inefficient strategies adopted for certain stimulus types or due to limitations in the perception of the stimuli (e.g. limits to the processing of peripheral stimuli).

Methods

Participants

10 observers (3 male) participated in the experiment including one author (JF). All observers had normal or corrected to normal vision and gave informed consent according to the declaration of Helsinki. All methods were approved by the ethics board at Queen Mary University of London.

Equivalent Noise Method

The two equivalent noise parameters are estimated by measuring *observer-noise* (specifically, their uncertainty on their estimate of the mean) as a function of changing *external noise*. The relationship between these data is described in equation 1. Observer noise is the sum of the internal and external sources of noise, divided by the number of samples used.

Equation 1: $\sigma_{obs}^2 = \frac{\sigma_{int}^2 + \sigma_{ext}^2}{n_{samp}}$

Where σ_{obs}^2 is the observer's discrimination threshold, σ_{int}^2 their internal noise, σ_{ext}^2 the added external noise and n_{samp} the effective number of samples used to estimate the mean.

In our experiment we quantified observer-noise by estimating their threshold for discriminating whether a group of faces is looking on average to the left or right of direct as our measure of observed noise. Thresholds were determined using a method of constant stimuli (MOCS). Observers are presented with ensembles whose mean offset is either to the left or right of direct-gaze and are required to indicate (reporting "left" or "right") the mean direction of gaze (or head direction) of the ensemble. By measuring performance repeatedly for a fixed number of offsets we can fit a psychometric function to each observer's performance (proportion of trials identified as rightward) and from this fit estimate each observer's discrimination threshold for different levels of external noise. The standard deviation of the normal distribution from which the gaze deviation or head rotation of each face is drawn corresponds to the external noise. At low external noise levels (narrow standard deviation), the faces will all be looking in approximately the same direction so the observer is limited by how well they can estimate the direction of any individual face (internal noise). When the external noise is high the faces will be looking in dissimilar directions, the external noise exceeds the internal noise, so observers will now be limited by the number of samples they are able to average (Fig 1a). By measuring discrimination thresholds at a range of external noise levels, we are able to fit a function to the data using equation 1 to obtain estimates for each observer's internal noise and effective sample size, for each stimulus type and presentation condition (e.g. Fig 2).

Stimuli

Sets of 8 gaze-deviations or head rotations were generated for the spatial and sequential averaging conditions. The individual gaze deviation stimuli were generated by first randomly choosing a facial identity from a set of four synthetic faces (2 male, 2 female) created using FaceGen (Singular Inversions 2016) software. The eyes were replaced with greyscale eye stimuli created in MatLab allowing for precise manipulation of gaze offset. To create the individual head rotation stimuli, the same four synthetic faces were loaded into Poser (Smith Micro 2016), a 3D model manipulation tool, and dark glasses were added to remove any cues from the gaze direction Figure 1. Using the software, we exported 1800 frames of an animation of each head rotating between 90° leftwards and 90° rightwards, producing stimuli with steps of 0.1° of head-rotation. All faces were then scaled so that they would subtend 4x4 degrees of visual angle during the experiment.

Individual face stimuli were combined to form spatial and sequential ensembles. For both the gaze deviation and head rotation stimuli the offset of each face was drawn from a normal distribution. The mean of this distribution was determined by the MOCS offset value for the given trial and the standard deviation was determined by the external noise level being tested. In the spatial condition, faces were presented simultaneously with each face randomly positioned within a 12.5 degree radius from the centre of the screen, such that no faces were overlapping (e.g. Fig 1a). For the sequential stimuli, faces were presented serially for 200ms each, separated by 200ms of a grey screen (e.g. Fig 1b). A small jitter (randomly chosen between up to 1° in all directions) was applied to the position of each face in the sequence to avoid any apparent motion effects. A lighter background grey colour was used for the head rotation stimuli, so that the edges of the faces were clearly defined.

Stimuli were presented on an Electron Blue CRT monitor (screen size 30x40cm) with a spatial resolution of 1600x1200 pixels operating at a frame rate of 85 Hz.

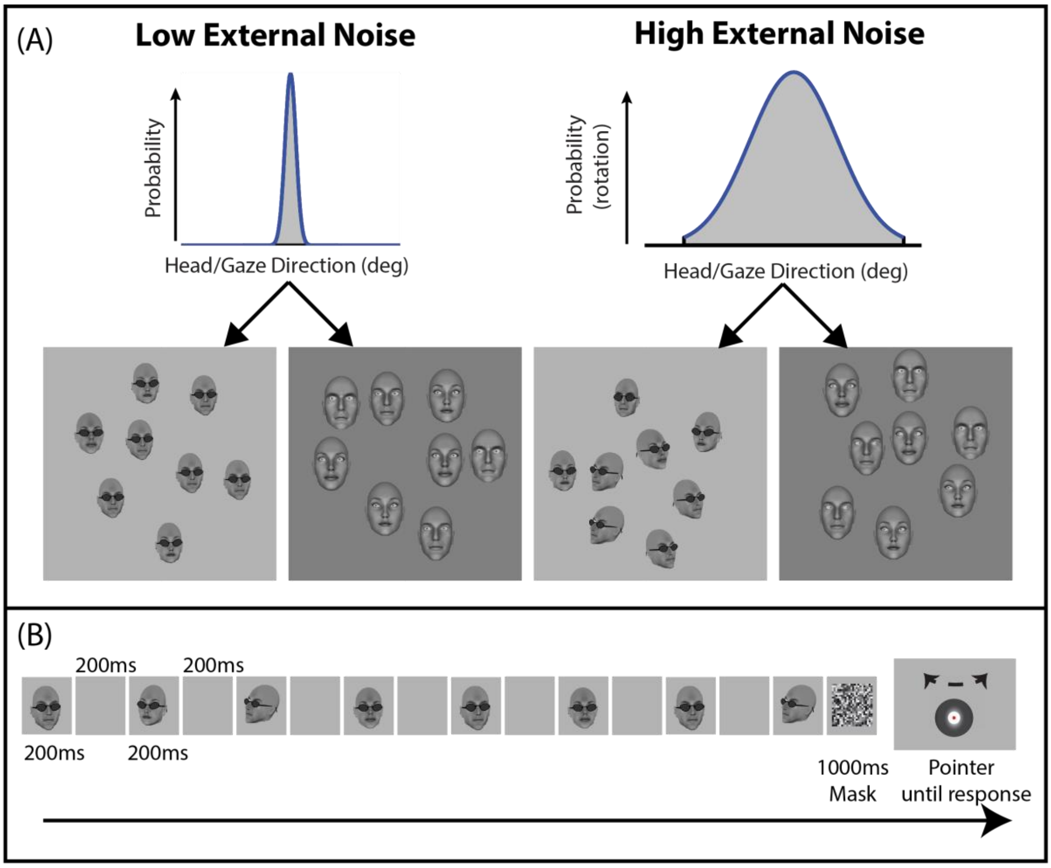

Figure 4:1: Examples of the stimuli presented in the four conditions

. (A) Two normal distributions from which the direction of each gaze or head rotation stimuli could be drawn. Below are examples of corresponding head rotation and gaze deviation stimuli that would be generated from these distributions, one with low external noise (all faces looking in the same direction) and one with high external noise (all faces looking in different directions). (B) A schematic depiction of a sequential head rotation. Faces are presented in a sequence with 200ms blank intervals between, followed by a noise mask and then a 3D pointer to indicate the average direction.

Procedure

Four sets of EQN parameter pairs were obtained for each of the four stimulus type/presentation type combinations (“Spatial Gaze”, “Spatial Head”, “Sequential Gaze” & “Sequential Head”). For each EQN function, thresholds were obtained at 6 levels of external noise. The standard deviations of the normal distributions were 0.5,2,4,8,16 & 24 degrees for gaze deviation and 0.5,2,4,8,16 & 32 degrees for head rotation. We set the highest noise level for the gaze deviation stimuli to 24 deg to avoid generating stimuli that exceeded the physical limits of human gaze (i.e. gaze offsets >60°). Two blocks of 80 trials were collected for each external noise level. Blocks included 10 repeats of the 8 mean offsets in a random order, producing a total of 160 trials. The mean offset values presented within any block depended on the external noise level of the block (to ensure even sampling of the psychometric function across conditions). For the gaze stimuli, noise levels below 5° standard deviation used offsets of [-15°, -6°, -3°, -1°, 1°, 3°, 6°, 15°] from zero and above 5° used offsets of [-20°, -10°, -5°, -1°, 1°, 5°, 10°, 20°]. For the head stimuli, three offset ranges were used: below 5° SD [-6°, -2°, -1°, -0.5°, 0.5°, 1°, 2°, 6°] , for SD=8 [-15°, -6°, -3°, -1°, 1°, 3°, 6°, 15°] and for SDs above 8 [-30°, -10°, -5°, -1°, 1°, 5°, 10°, 30°] . Blocks for a single condition were collected in approximately hour long sessions with a randomised order of external noise levels.

Experimental control and stimulus presentation were controlled in Matlab (Mathworks Ltd) using Psychtoolbox (Brainard, 1997). In the spatial blocks the 8 faces were presented simultaneously for 1600ms . In the sequential blocks each face was presented for 200ms separated by 200ms of a blank screen followed by a 1000ms noise mask. In both presentation conditions, the stimulus was followed by a “3D” response pointer which could be rotated with the mouse. The observer rotated the pointer and clicked to indicate when it was pointing in the mean gaze direction/ head rotation of the set of faces. No feedback was given and the next trial commenced 200ms following the response.

Threshold and equivalent noise Fitting

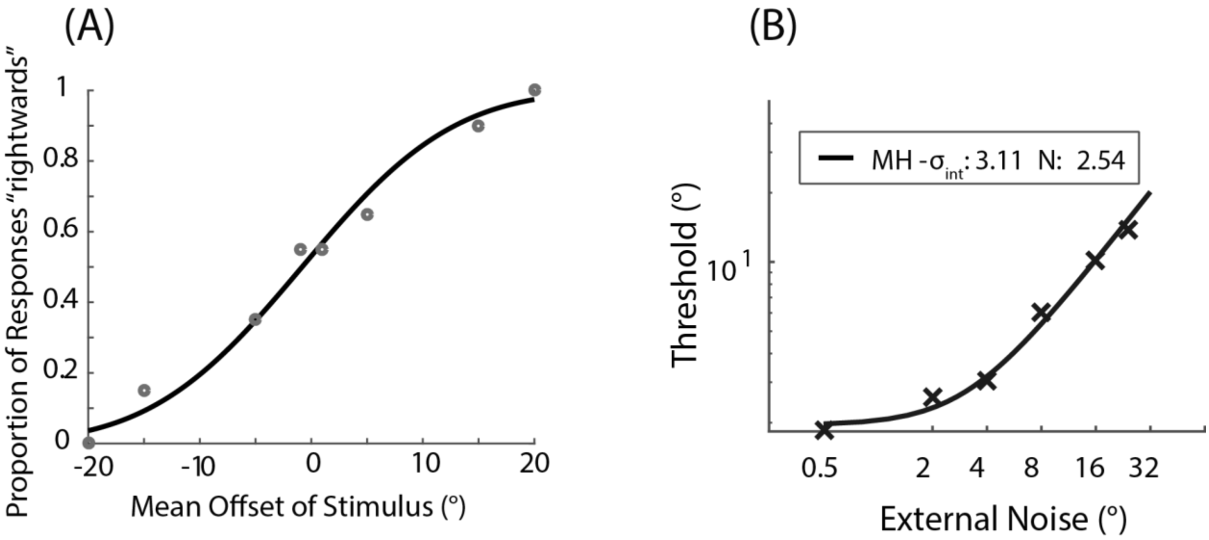

Figure 4:2 Example threshold and equivalent noise fits.

(a) The proportion of times the participant responded "rightwards" to a given mean offset is plotted (blue circles) against the mean offset of the ensemble. A cumulative Gaussian function is fit to these data (black line) the slope of which is the observers' discrimination threshold (which quantifies their uncertainty about the ensemble-mean). (b) An EQN plot. The thresholds ("x" symbols) are plot against the corresponding external noise level. The black line is the fit from the model described in the text. The inset shows estimated internal noise (/sigma) and effective sample size (N).

Observers' response were converted to 1 (positive) or -1 (negative) to indicate an overall leftwards or rightwards response respectively. Data from two separate runs for each participant were combined, giving 20 repeats at 8 different offset levels. A cumulative Gaussian function was fit to the proportion of times the participant responded "rightwards" for each mean offset direction (Fig 2a) using a maximum likelihood method. The standard deviation of this cumulative Gaussian function was taken as the discrimination threshold for the participant at a set level of external noise.

Discrimination thresholds quantify observer-noise and (for a single participant and single stimulus/presentation combination) are plot against the external noise levels (Fig 2b). The equivalent noise function (equation 1) was then fit to these threshold values (solid line, Fig 2b), yielding estimates of internal noise and effective sample size.

Results

The results for the EQN analysis are summarised in Figure 3a. A 2x2 (Gaze x Head, Spatial x Sequential) repeated measures ANOVA was conducted for each of the two EQN parameters, internal noise and sampling efficiency. For internal noise, there was a main effect of stimulus type $F(1,9) = 22.1$, $p=.001$. Pairwise comparisons revealed that gaze deviation averaging was associated with significantly more internal noise than head rotation averaging ($p=.001$). There was no main effect of the presentation conditions $F(1,9) = 1.49$, $p=.25$, nor a significant interaction $F(1,9) = 2.19$, $p=.17$. The ANOVA for effective sample size revealed no significant main effects for stimulus type $F(1,9) = .588$, $p=.46$ or presentation type $F(1,9) = 2.58$, $p=.12$. There was however a significant interaction $F(1,9) = 20.4$, $p=.001$. Paired sample t-tests show that for the spatial presentation, head rotation was associated with a significantly greater effective sample size (M=3.8) than for gaze deviation (M=2.5), $t(9) = 2.9$, $p= .018$. For the sequential presentation, there was no significant difference between the two stimulus conditions $t(9) = 1.6$, $p= .138$.

Taken together, these results show that observers are more uncertain about the direction of individual elements in gaze deviation stimuli compared to head rotation stimuli of the same size and presentation duration. The results for effective sample size suggest that observers can use more elements to average groups of head rotation than gaze deviation stimuli, but only when faces are presented simultaneously in the spatial conditions.

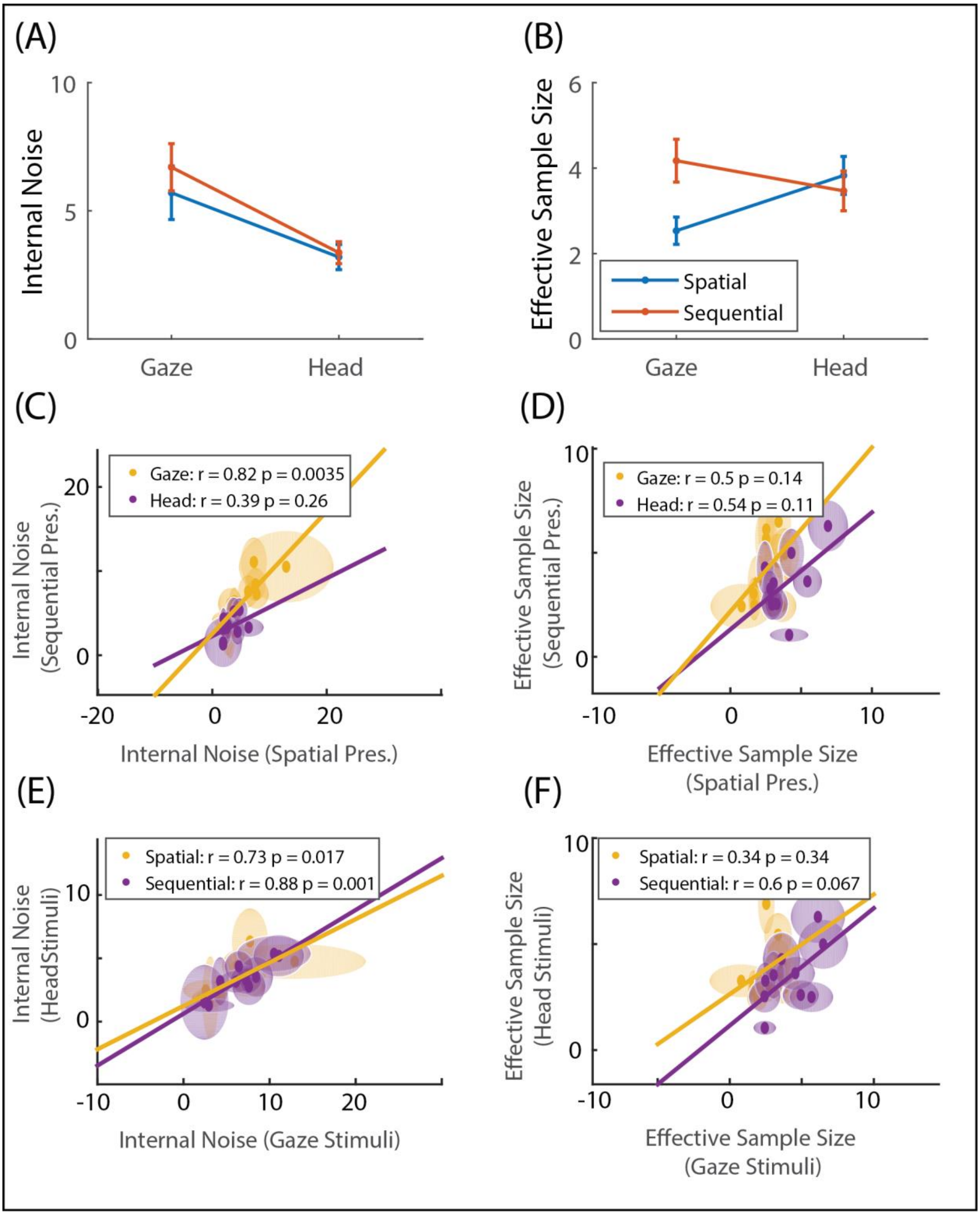

Figure 4:3 Summary of equivalent noise and correlation results.

– (A,B) Two plots showing the mean equivalent noise parameters (internal noise and effective sample size) for the two stimulus types (x-axis) for each of the presentation types (red/blue lines). Error bars show +/- one standard error of the mean. (C,D) Scatter plots for the relationship between the two presentation conditions for each stimulus type (yellow = gaze, purple = head). Data from each individual

is represented by the individual points and a best fit correlation line is drawn through the data. Figure legend shows the correlation coefficient *r* and significance of the correlation *p*. (E,F) As in (C,D) but now for the relationship between the two stimulus conditions for each presentation type.

Pearson's product moment correlation analyses were performed for each EQN parameter, for the two stimulus conditions and the two presentation conditions (Fig 3b,c). For internal noise a significant positive correlation (r=0.82, p=.003) was found between the two presentation conditions for gaze stimuli but not for head rotation. Across both presentation types, a significant positive correlation was found between the two stimulus types (spatial: r=0.73, p=.003, sequential: r=0.88, p=.003). For effective sample size, no significant correlations were found, though there was a borderline significant (p=0.067) relationship between the two stimulus types in the sequential condition.

These results suggest that observers who have high internal noise for gaze stimuli will also have high internal noise for head stimuli for either presentation type. Similarly, observers who have high internal noise for spatially distributed gaze stimuli also have high internal noise for sequentially presented gaze stimuli. The borderline significant correlation may suggest that observers who have high effective sample size for gaze will also have high effective sample size for heads, but only when both are presented sequentially.

Discussion

Using an equivalent noise procedure we have compared averaging of head rotation and gaze deviation, for stimuli matched in size and presentation duration over both space and time. From these data, we estimated observers' internal noise (the uncertainty of an individual in estimating the direction of a single face from the array) and effective sample size (the number of samples they are able to effectively average) for ensembles of eight face stimuli. We report that head rotation is averaged with a greater effective sample size than gaze deviation in the spatial condition but not in the sequential condition. In both presentation conditions, gaze deviation judgements were associated with more internal noise than head rotation. A correlation analysis revealed a relationship between observers' internal noise for the two stimulus types, and between the two presentation types for gaze stimuli. A possible relationship was also found between the effective sample sizes of the two stimulus types when sequentially presented, though this was only weakly significant.

The results for the internal noise estimates are consistent with our previous results for gaze deviation averaging (Florey et al., 2016). This difference cannot be attributed to the peripheral presentation of the faces, since we report similar results using both spatial and sequential presentations, with the latter presenting all faces at the fovea. Although previous results have suggested similar precision in estimating gaze and head direction (Loomis et al., 2008), we find that for an averaging task, head rotation elements are processed with less uncertainty. This may be because attention is necessarily spread over either space or time, and that gaze deviation requires more focused attention to process with a high level of precision. Consistent with our previous findings, this suggests that when judging the direction of a crowds' attention, head rotation is used for rapidly summarising the direction of attention, then gaze can be used to judge the interest of any individual within the crowd.

It is somewhat surprising that there was a significant difference in the effective sample size between the two stimulus conditions for the spatial presentation. Although we had previously found this difference, for briefly presented stimuli (300ms), we expected that difference to disappear (or be largely reduced) here. This is because (a) observers would be able to make multiple saccades in the 1.6s that the stimuli were presented for and (b) the stimuli were now larger, so that the limitations on peripherally processing gaze stimuli would be reduced. A possible explanation for our result is that observers may not efficiently make saccades across the groups of faces (e.g. rather than foveate a single face they saccade in between two faces). This would mean that their average would still be limited by the fact that some faces fall in their periphery and so were not used in the average computation. This difference between the two types of cue (head / gaze) is not present in the sequential presentation condition, suggesting that this is not due to the specific stimulus per se (i.e. that observers are poor at integrating gaze deviation signals), but rather to the spatial distribution of the elements in an array. This has important implications for averaging research that compares different types of stimuli since the distribution of the elements as well as their peripheral perceptibility must be carefully controlled to avoid effects simply being the result of limited peripheral processing.

The correlation results for internal noise provide an interesting insight into what limits individuals in their averaging performance. We find that internal noise is highly correlated across the two types of stimuli. This suggests that there is a source of noise that may be generic to these two different stimulus types. Solomon (2010) suggests that there are two sources of noise that affect the processing of an ensemble, one that acts on the individual elements and one that acts on the process of computing the mean estimate generated from these estimates. The differences we find in internal noise between the two stimulus conditions suggest that this early source of noise is independent, possibly arising at a later stage, such as the superior temporal sulcus where cells are known to specifically encode head and gaze information. The correlation results suggest that within an individual, the later source of noise may be agnostic to the stimulus domain and may act on the processing of any ensemble. Alternatively, this source of noise may arise at an early level (such as the primary visual cortex), limiting both types of stimuli as a result of noisy processing of low level features.

The finding that effective sample size does not correlate between the two stimulus types in the spatial condition is potentially inconsistent with Haberman et al. (2015) who report that individuals' averaging performance was correlated between two different face based tasks (face emotion and identity). It may be that their correlation came as a result of a shared source of internal noise as opposed to similar sampling efficiency between tasks. Alternatively, it may be that the reason we do not see a correlation is because individuals sampling efficiency could be confounded by individual differences in peripheral visual perception. Observers may have the same limits on their sampling efficiency between the two stimulus types, but because they are also limited by their peripheral perception of gaze, this correlation does not become apparent. The fact that for sequential averaging, the two stimulus types were weakly significantly correlated but this was not the case for spatial averaging provides some support for this interpretation.

Our finding that there is no relationship between sampling efficiency between the two presentation types, suggests there is no generic limit on the integration of multiple samples independent of the way they are presented. Most likely independent limits, such as spread of attention and sampling strategy in spatial ensembles; and short term memory and temporal biases in sequential ensembles; have a greater influence on sampling efficiency (and performance) than any generic limit imposed by a single averaging mechanism.

Clearly there are many limits that must be considered in averaging tasks; here we address some of the issues for simultaneously presented stimuli and suggest that care must be taken to ensure that any differences in peripheral perception are controlled for, even when using long presentation durations. In addition, when considering individual differences in averaging ability, it is important to consider whether performance is being limited by internal noise or sampling efficiency, as the two can vary independently.

Chapter 5 The Effect of Weighting Strategy on Sequential Averaging Performance

Introduction

People are able to extract summary statistics from ensembles of items (also known as ensemble coding) presented either across space or in a temporal sequence. This has been demonstrated using a wide variety of stimuli, from low level properties like size (Ariely, 2001; Gorea et al., 2014), orientation (Dakin & Watt, 1997; Dakin, 2001; Solomon, 2010) and motion (Dakin et al., 2005), to high level stimuli such as facial emotion (Haberman et al., 2009) and gaze direction (Sweeny & Whitney, 2014). In previous studies (and chapter 3), it has been shown that observers are not able to efficiently combine all the information presented to them to form an average (e.g. Dakin 2001, Florey et al. 2016). Instead, they perform as if they are only using a sub-set of the stimuli available. For example, in chapter 3, results indicated that observers always used fewer than the eight faces presented to judge the average gaze deviation or head rotation of a crowd; both when presented simultaneously (in space) and sequentially (in time).

The reason for this inefficient pooling is not clearly understood but there is some evidence that the strategy the observer employs can influence their averaging performance (e.g. Florey et al., 2016; de Gardelle & Summerfield, 2011). Strategy here refers to the biases, either conscious or unconscious, that the observer adopts when processing stimuli (or features) to average. In chapter 2 it was found that observers were biased to use elements that appear in the centre of a spatially distributed group and other studies have shown biases toward using stimuli that are close to the mean of the total set (de Gardelle & Summerfield, 2011). It is not clear to what extent these biases are conscious or automatic, though evidence clearly supports the idea that people do not average information in the way that an ideal observer would; that is, by precisely estimating the value of every element presented to them and computing a perfect linear average of all elements. In chapter 2, the reasons for limited spatial averaging were investigated; however, the possible reasons for inefficient averaging of temporal sequences was not addressed. Here we aim to investigate the effect of temporal biases (e.g. whether people weight some samples more than others across a sequence) on averaging performance and to see if changing the properties of a stimulus sequence can shift observers' strategy or induce changes in averaging performance.

The image that falls on a persons' retina is constantly changing, both through variations occurring in the environment (e.g. a face with shifting expressions) and through eye and head movements. Various studies have investigated how we are able to process this complex sequential information to form useful summary statistics. Haberman, Harp and Whitney (2009) found that observers could more accurately report the mean facial emotion from a rapidly presented sequence of faces than for any individual face taken from the set. Similar results have been obtained for size (Oriet & Brand, 2013)and auditory pitch (Albrecht et al., 2012). Although these studies demonstrate that there is clearly a mechanism for temporal averaging, they do not reveal how it operates or what temporal limits it may have.

Averaging, whether over space or time, has regularly been shown to be inefficient; observers do not perfectly integrate information from every sample they are presented with to calculate a mean. This has been shown for spatial averaging of orientation (Dakin, 2001) motion (Dakin et al., 2005) and social cues over space (Florey et al., 2016). Possible explanations for limited spatial processing include observer's inability to rapidly distribute their attention across an entire display (Attarha & Moore, 2015; Brand et al., 2012) and observers limited processing of visual elements in their periphery. Florey et al. (chapter 2) showed that observers were more likely to use elements in the centre of an array of faces than those in the periphery, though (Wolfe et al., 2015) have shown that observers are able to perform averaging tasks using only peripheral elements. These explanations go some way to explaining why observers are unable to efficiently combine information from a spatially distributed group of items, though these limitations should not affect the processing of sequential ensembles (since both attention and fixation can be focused on a single element at a time). Therefore there must be some other limits that prevent observers from perfectly integrating sequential averages.

Averaging of stimuli over time has been shown to be sub-optimal. For example, Gorea et al. (2014) found that when averaging the size of dots presented sequentially at fixation, observers were only able to use 4-6 elements from a set of 8. A similar result was found for averaging of gaze deviation and head rotation (Florey et al., 2016/Chapter 3). Observers were presented with a sequence of eight faces whose gaze deviation or head rotation were drawn from a normal distribution. When the standard deviation of this distribution was low (i.e. faces were looking in approximately the same direction) observers could accurately report the mean of the set. When the standard deviation was high, so that faces were all looking in different directions and judging the average was much harder, observers were no longer able to accurately report the mean of the entire set. Equivalent noise analysis of this data (see chapter 2/3 methods) was used to show that observers were significantly worse than an ideal observer who used all the samples available to them. Instead they performed as if they were using only a sub-set of the elements available.

There are a number of potential biases in the perception and integration of sequence ensembles that may account for the limited processing. Hubert-Wallander & Boynton (2015) have recently shown that observers do not weight all the items in a temporal sequence equally when they calculate the average. Participants completed similar sequential averaging tasks for a number of different stimulus types (position and size of dots, motion direction of a moving dot field and facial expression). For each stimulus type, participants were presented with sequences of 10 (size/position) or 8 (motion and faces) stimuli and asked to report the mean property of that set. For example, in the size task, observers were presented with 10 circles of different sizes and then asked to adjust a "response" dot to match the average size of the dots presented in the sequence. For all stimulus types, observers responded along a continuous scale, giving their estimate of the mean of the ensemble. In order to determine the weighting each element in the sequence was given for the mean calculation, the authors used a linear regression analysis. In their regression model, the observer's predicted response on each trial was the sum of the 8 (or 10) elements presented, each multiplied by a different weighting. The best fitting weighting parameter for each position in the sequence gave an indication of how much that position was weighted into the observer's average. Surprisingly, they report that different patterns of weightings emerged for different stimulus types rather than a generic mechanism effect across stimulus type. For example, averaging of face emotion was associated with a recency effect, where elements at the end of the sequence were weighted more strongly than those at the start. Gorea et al. (2014) report a similar recency effect in a size averaging task. Conversely, in a positional averaging task, elements at the start of a sequence were shown to have higher weightings; an effect known as primacy. Unequally weighting all items in a sequence produces sub-optimal averaging and may be responsible for the apparent sub-sampling that has been previously observed.

There are other temporal biases that may affect temporal averaging. For example, in many studies serial dependency effects have been observed. In serial dependency, the perception of an item is biased towards the (value of the) item preceding it. These dependency effects have been shown in sequences of oriented Gabors as well as face emotion stimuli (Fischer & Whitney, 2014; Liberman, Fischer, & Whitney, 2014). This account of temporal biases is consistent with the primacy effects found by Hubert-Wallander and Boynton (2015) for dot position, as it implies that the first item in a set will influence the second, and so on, so that every element is influenced by the earlier items. This is of course inconsistent with the more commonly observed recency effects (e.g. Hubert-Wallander and Boynton, 2015, for faces and size and Gorea et al., 2014, for size), suggesting that serial dependency effects may not be present for all stimulus types, or that the effect is sufficiently weak that it is swamped by stronger effects in sequential averaging.

De Gardelle and Summerfield (2011) have suggested that averaging is biased towards items that are close to the mean of the distribution to be averaged, so called “robust averaging”. This was initially demonstrated in spatially distributed arrays of colour stimuli and more recently used to model sequential integration of orientation information (Cheadle et al.., 2014). Cheadle et al. found a similar recency effect to that which has been reported before, but also found that their data was best modelled by an integrative mechanism which increases the weighting of elements that are consistent to those preceding it. They suggest that it is efficient to weight information more highly if it is expected. That is to say, if it is consistent with the estimate of the population mean that has been established to that point. These results again suggest that observers do not weight the elements in a sequential averaging task equally and may account for inefficient averaging performance.

As well as inherent biases in averaging strategy, there is also evidence that observers can adapt their strategy depending on properties of stimuli. Juni, Gureckis, and Maloney (Juni et al., 2012) presented observers with sequential cues to the position of a target and then observers were required to estimate the location of the target. Crucially, the variance of the normal distribution from which the position of each cue was drawn was dependent on the position of the cue in the sequence. In one condition, the distribution for each cue had decreasing variance such that the cues became more precise across the sequence. In the other condition the distributions had increasing variance so the cues became increasingly less precise over the course of the sequence. Participants were rewarded for correct responses. They found that over multiple blocks of trials, participants improved their performance by changing how they weighted the items in the sequence, such that they increased the weighting of more precise elements. This happened regardless of whether the precision of each item increased or decreased across the sequence. This suggests that observers are sensitive to the global properties of the sequence (i.e. the precision of each element's representation of the mean) and that they can adapt their strategy to exploit these properties.

In the present study, we first analyse data from chapter 3 to see if we replicate the recency biases found in other sequential data and to see if the amount of bias observed can predict the observers' accuracy in averaging. We will then test the flexibility of observers averaging of head rotation in two stimulus conditions; one where the faces in the sequence to average are presented in a random order and one where the stimuli are sorted so those close to the mean are presented towards the end of the sequence. The aims of this experiment are threefold: (1) to determine if the amount of bias in observers' responses correlates with their averaging precision? If elements are weighted more equally then the average derived from the sequence should be more precise than a more biased weighting of elements. (2) To determine whether altering the properties of the stimuli alters the observer's strategy. For example, if stimuli are consistently tailored towards a certain type of strategy (e.g. recency) will observers alter their strategy to suit this bias (i.e. upweighting later items when they are closer to the mean than early items), or do observers stick rigidly to their original strategy? (3) to determine if we can exploit these biases to improve averaging performance in observers?

Regression analysis of sequential head and gaze deviation averaging

In chapter 3 data was collected from 10 participants averaging sequences of both gaze deviation and head rotation. Participants used a 3D rotating pointer to indicate the average gaze deviation or head rotation of a set of 8 faces (for full methods see chapter 3). This data can be analysed using a linear regression in a similar way to previous sequential averaging studies**.** This will allow us to estimate the weighting of each gaze or head rotation stimuli in the observers mean calculation. Given Hubert-Wallander and Boynton's (2015) found primacy effects when judging the mean position of a dot, we may expect that gaze deviation averaging (which requires observers to average the mean position of the iris in the sclera), would produce the same pattern of results. However, in the same study, the authors find a recency effect for a face emotion averaging task, so it may be that we find a recency effect for our data if all face related tasks share this same bias.

The data from chapter three provide a measure of averaging performance for each participant in the form of *effective sample size* (ESS - see chapter 3 methods for full details) that gives an estimate of how efficiently observers combine information from the samples. The more samples an observer is combining the higher their ESS will be, up to a maximum of n out of n samples (in this case 8). An observer who is using all the available samples efficiently should not be biased to weight those later or earlier more highly than any other, so we may expect that observers with high ESSs would display less varied weightings than those with lower ESSs.

Regression Modelling

In order to estimate the weightings of each element in a sequence, a simple weighted sum model was used, similar to Juni et al. (2012) and Hubert-Wallander and Boynton (2015). The mean perceived average (as measured by the pointer response the participant gave – *R in degrees*) is the sum of each head rotation angle (x_i) multiplied by a weighting factor applied to that position (w_i) plus a constant (c) (see equation 1). Simple linear regression was used to obtain the least square best fitting estimates for the weighting of each position in the sequence. This was done separately for each participant (fig 2a) and, for the different order conditions.

Equation 1:

$$R = \sum_{i=1}^{8} w_i x_i + c$$

To check if each observer's fitted model was significantly better than a model where all elements are equally weighted, an F-ratio test was conducted for each observer, for each stimulus type. This takes the ratio of the residuals of two models, in this case an observer's fitted model and the an equally weighted model, to produce an *F* statistic which can be used to determine if one model provides a significantly better fit than another.

The regression analysis was applied to the data from chapter 3 for each participant on the combined data from the three highest external noise levels, giving 480 trials for each participant and stimulus type combination. The lower external noise levels were excluded as the task was very easy and the stimuli were all facing in approximately the same direction so observers decisions are not informative as to which elements they were using.

First, to look at ”position in sequence” effects like recency and primacy, the linear regression was applied with *i* being the position of each face in the sequence (i.e. 1-8 from first to last presented). In a second version of the analysis, the values presented to the observer on each trial were sorted by their distance to the mean (i.e. 1-8 with 1 being the stimuli whose value was closest to the mean of that trial and 8 being the furthest from the mean) before applying the regression analysis. This allowed us to look for “robust averaging” effects, where stimuli that are close to the mean of a set are expected to be up-weighted.

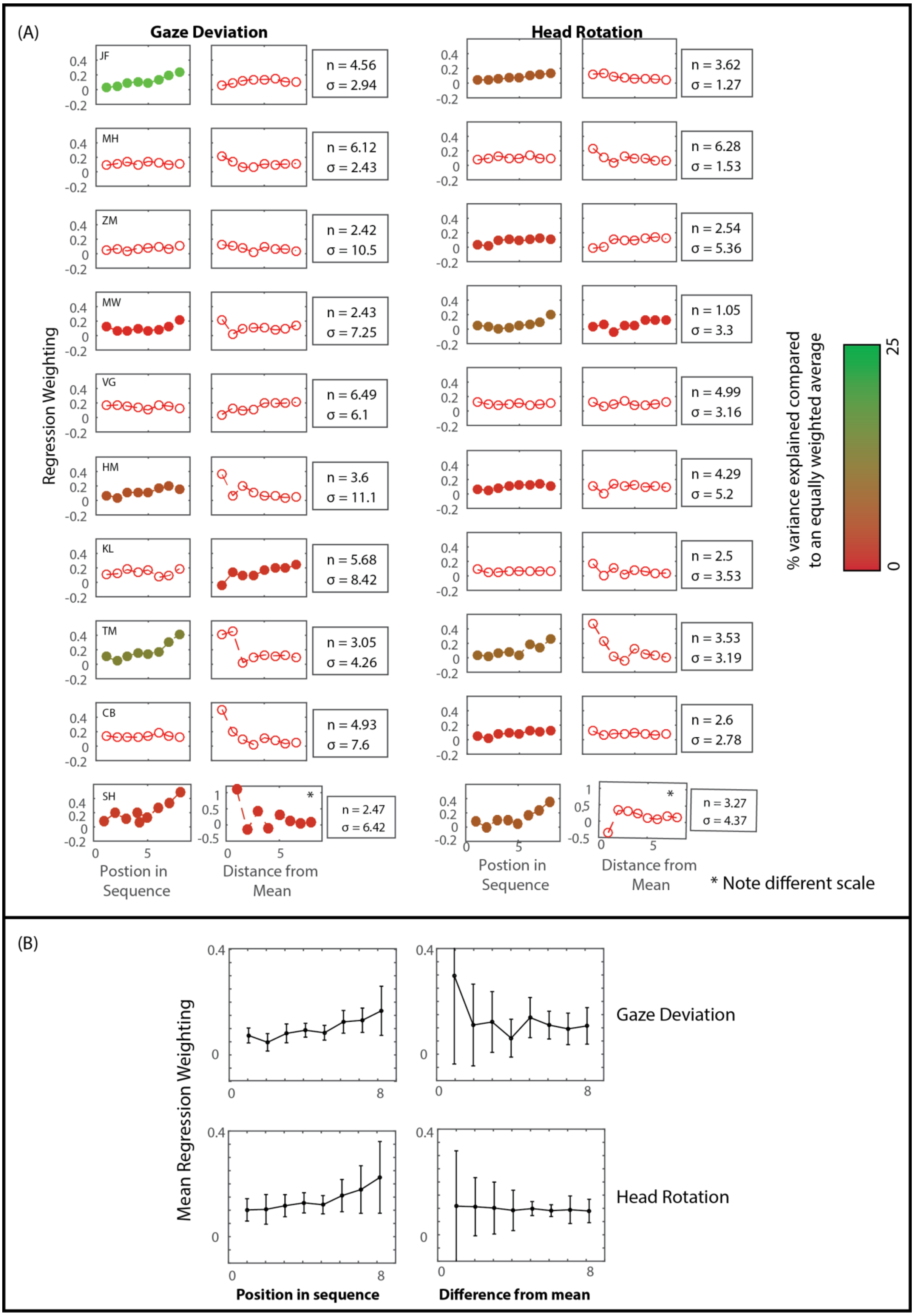
(A)
Gaze Deviation
Head Rotation
Regression Weighting
JF
MH
ZM
MW
VG
HM
KL
TM
CB
SH
n = 4.56
σ = 2.94
n = 3.62
σ = 1.27
n = 6.12
σ = 2.43
n = 6.28
σ = 1.53
n = 2.42
σ = 10.5
n = 2.54
σ = 5.36
n = 2.43
σ = 7.25
n = 1.05
σ = 3.3
n = 6.49
σ = 6.1
n = 4.99
σ = 3.16
n = 3.6
σ = 11.1
n = 4.29
σ = 5.2
n = 5.68
σ = 8.42
n = 2.5
σ = 3.53
n = 3.05
σ = 4.26
n = 3.53
σ = 3.19
n = 4.93
σ = 7.6
n = 2.6
σ = 2.78
n = 2.47
σ = 6.42
n = 3.27
σ = 4.37
Postion in Sequence
Distance from Mean
* Note different scale
% variance explained compared to an equally weighted average
25
0
(B)
Mean Regression Weighting
Gaze Deviation
Head Rotation
Position in sequence
Difference from mean

Figure 5:1 –Summary of the regression analysis on data from chapter three

. (A) Each coloured plot shows the weightings derived from the linear regression analysis for each element in the sequence and for each stimulus type (gaze/head), each participant and each regression condition (sorted by order presented or by distance from the population mean). The colour of the line shows the percentage of extra variance explained by the regression model compared to one with equally weighted elements. Filled symbols show situations where the regression fit is significantly better than an equally weighted one, according to an F-ratio test. Equivalent noise parameters (internal noise and effective sample size) for each participant are shown next to the individual plots. (B) Mean weightings for each condition are shown in black plots, error bars show +/- one standard deviation.

Results

The results of this analysis are summarised in figure 1. It is immediately clear that there are individual differences in the way that people weight elements into their average, reflected in the range of patterns from each participant (e.g. compare JF to VG) and the wide error bars on the mean weightings plots. Although some observers show no biases in their weighting of the elements for the "position in sequence" analysis, there is an overall recency effect. Repeated measures ANOVAs were carried out on the mean weightings for the two analysis types (order sorted and mean sorted) and two stimulus types (heads and gaze). In the order sort condition, both stimulus types ANOVAs were significant, gaze: $F(1.68,15.1) = 5.27$, $p = 0.02$; head: $F(1.61,14.5) = 7.52$, $p = 0.08$. In both cases we find a recency effect, where items later in the sequence were weighted more heavily than those at the start. Neither ANOVA was significant for the mean sorted condition.

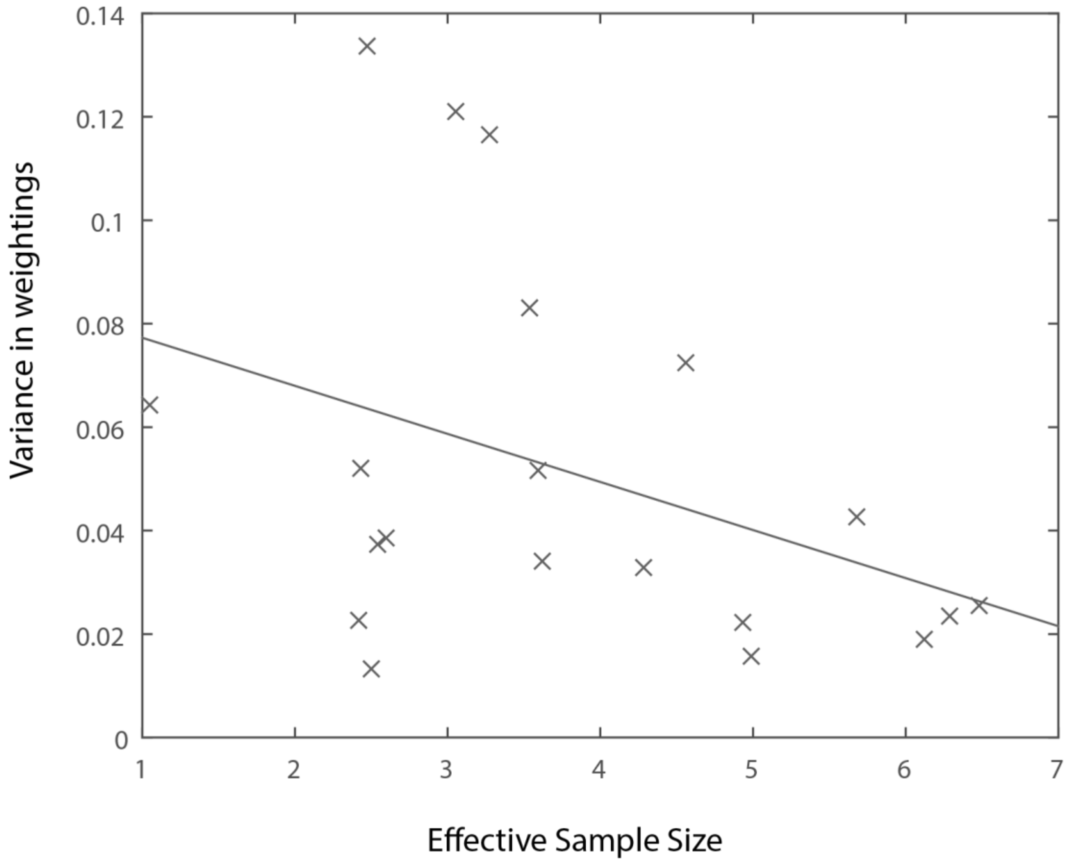

Figure 5:2 –Correlation between the effective sample size for a participant and the amount of variance in their regression weightings

. Data are combined from both gaze deviation and head rotation conditions.

To see if there is a relationship between an observers' effective sample size and the amount of variance in their weightings of each element, a correlation analysis was carried out between these two variables on the combined data for head rotation and gaze deviation. Although there does appear to be a small negative correlation between the two, where lower weighting variance is correlated with higher effective samples sizes, this relationship is not significant (r = -0.37, p = 0.09).

Discussion

For both stimulus types we find a recency effect, where elements that appear later in a sequence are weighted more heavily into an observer's perceived average than earlier items.

No difference was found between gaze deviation and head rotation averaging, suggesting observers employ the same mechanism for both types of social cue. This recency effect is consistent with the findings of Gorea et al. (2014) and with Hubert-Wallander and Boynton's (2015) results for size and face emotion averaging. These results suggest that the primacy bias found by Hubert-Wallander and Boynton (2015) for position averaging does not affect gaze averaging, even though judging gaze direction relies on perceiving the position of the iris. It is possible that a recency bias will always dominate in a face based task as the most important information in a social context is likely to be the most recent information; that is to say, even if there is a small primacy bias for position it may be swamped by a stronger recency bias.

The amount of variance between individuals suggests that observers' strategy can vary greatly even within the same task. Although we find an overall recency effect, some observers' weighted regression fits did not fit their data significantly better than an equally weighted model (filled symbols on figure 1). This could be the result of some observers adopting conscious strategies to intentionally use all the information they are given, while others are only driven by what may be a more automatic, recency bias. Future research into individual differences in averaging is needed to properly understand these effects.

No significant effects of "robust averaging" were found using the regression analysis. There does seem to be an effect in some individuals (e.g. SH), though overall the distance that the value of an element was from the mean of the set did not have an impact on how much the element was weighted in the average. It may be that there is some small effect of elements close to the mean being up weighted but this is much smaller than the recency effect that we observe here.

It was expected that observers who weighted items equally would have a higher ESS than those who exhibited larger biases; however, this was not the case. No significant correlation was found between these two variables, though there was a trend in the expected direction. This suggests that observers who are more efficiently integrating all the information available to them are not necessarily doing this by weighting all items equally. It seems that the reason for observers performing as if they are using only a sub-set of the available information is not because they are ignoring some elements in their calculation. What exactly causes this limit in averaging performance remains to be determined.

Experiment 2: Does biasing stimuli change averaging strategy?

Having found that the recency bias observed in experiment 1 does not correlate with sampling efficiency, it is unclear exactly what these apparent strategies tell us about actual averaging performance. Juni et al. (2012) showed that observers could shift their strategy depending on the properties of the stimulus sequence, so we wondered whether observers adapt their strategy if the stimuli are tailored to a certain type of strategy. For example, if stimuli are designed so that having a recency bias is advantageous, will observers who previously showed no bias, produce a recency effect? If not, do observers who already had a recency bias improve when averaging these tailored stimuli? A new data set was collected on a set of naïve participants to answer there two questions.

Methods

Participants

Participants were one author (JF) and 17 naïve observers from the undergraduate psychology program at Queen Mary University of London. All participants had normal or corrected to normal vision and all methods were approved by the university ethics board and complied with the declaration of Helsinki.

Stimuli

Stimuli consisted of sets of 8 faces with dark glasses covering the eyes, rotated to different directions. There were two stimulus conditions tested for each participant; (1) a random condition and (2) a recency sorted condition. In both cases the individual elements in the stimuli followed a similar creation. To generate each individual head rotation element, 3D head rotation stimuli were generated using a combination of FaceGen, Poser and Blender software (for full details see Florey et al. 2016). All faces were scaled so that they subtended 4x4 degrees of visual angle. Stimuli were presented on an Electron Blue CRT monitor (30x40cm) with a resolution of 1600x1200 pixels.

The rotation of each head in the sequence was drawn from a normal distribution with a standard deviation of 16°. In the recency sorted condition the randomly generated head rotations were sorted such that element position in the sequence was inversely correlated to the distance between the individual stimulus value and the mean (e.g. first element value was furthest from the mean and last element was closest) . In the random condition the faces were presented in a random order. In both conditions each face was presented for 200ms with a 200ms grey screen between each face. At the end of the sequence a noise mask was presented for 1000ms and then a pointer for response.

Procedure

On each trial, participants were presented with a sequence of 8 rotated faces followed by a 3D response pointer. Observers rotated this pointer with the mouse and clicked to indicate the mean direction that the faces had been looking in. No feedback was given.

There were 8 different mean directions [-30°, -10°, -5°, -1°, 1°, 5°, 10°, 30°] each presented 10 times per block. Each participant completed three blocks for each order condition in a random order, resulting in a total of 240 trials for each stimulus condition.

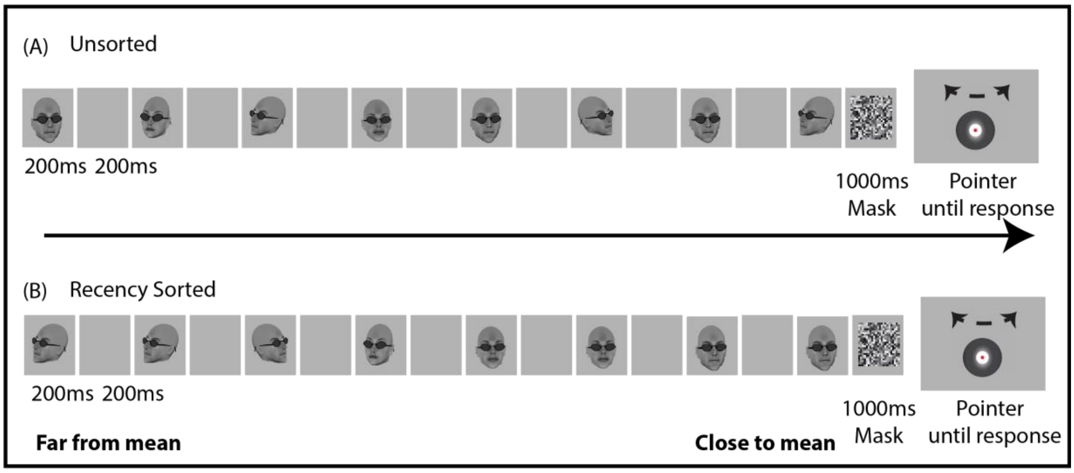

Figure 5:3– Example stimuli for the random (a) and sorted (b) condition

. A Sequence of 8 faces is presented followed by a noise mask and then a 3D response pointer. In the random condition the direction of each face is drawn from a normal distribution and presented in a random order. In the sorted condition faces are also randomly drawn from the same distribution but are then sorted so that those closest to the mean appear at the end of the sequence.

Results

The same regression analysis and F-ratio tests were used to estimate observers' weightings for each element. For 30/36 regression fits the differently weighted fits were significantly better than an equally weighted average (see figure 4a, filled symbols). The majority of participants show some recency effect, though there is clear variance in the size of this effect (e.g. compare ACM to GS). Regression weightings to the mean data reveal a small recency effect. One way 1x8 ANOVAs on each order condition revealed significant differences in the weightings for both random ($F(2.7,46.6) = 3.73$, $p = .02$) and sorted order conditions ($F(3,50.6) = 5.03$, $p = .004$). Visual inspection reveals that differences between the two conditions were minimal; observers appeared to have a single weighting profile which did not change notably between the two conditions.

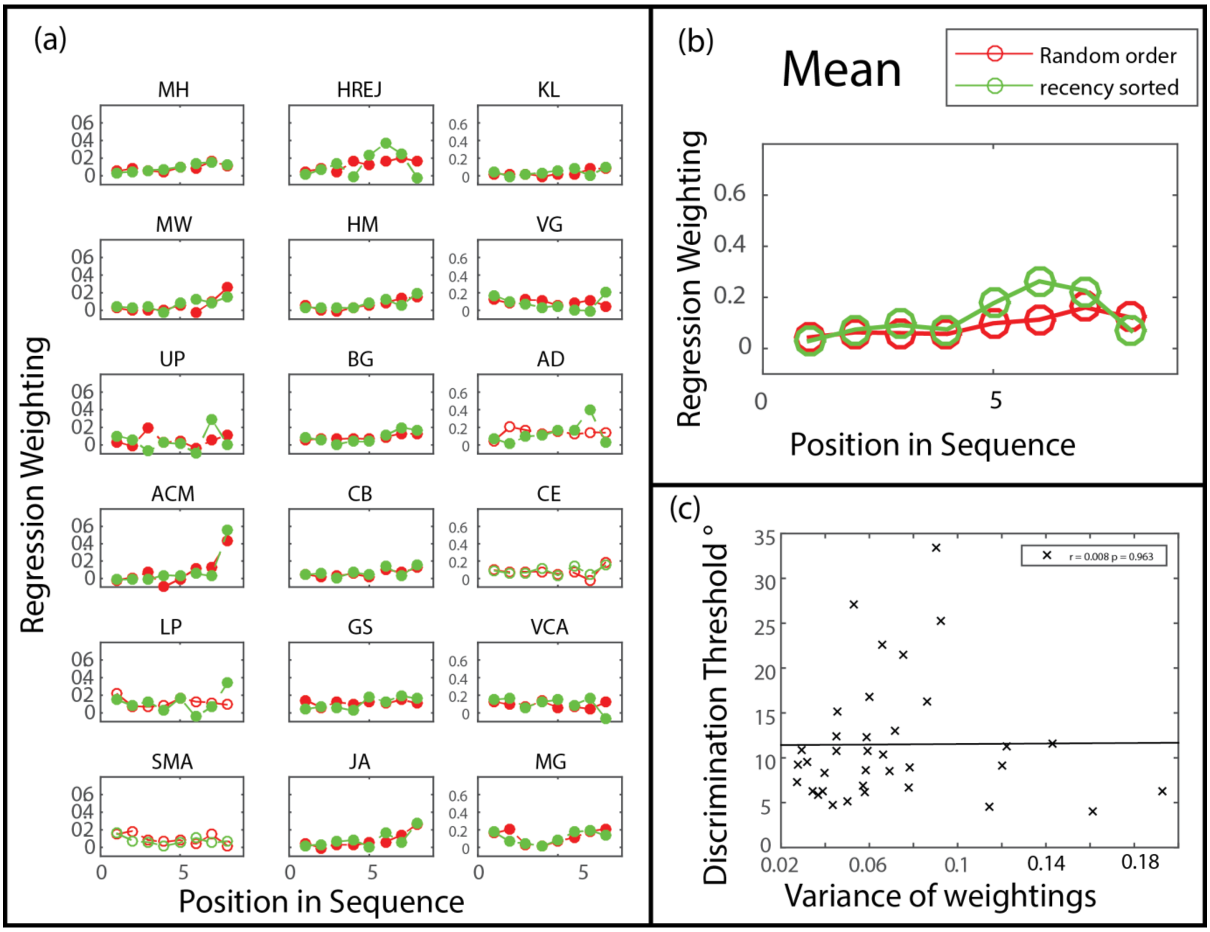

Figure 5:4– Summary of the regression analysis results

. (a) The regression weighting of each element in the sequence of 8 faces is plotted against the position of each element in the sequence from first to last. Each plot shows data from an individual observer and each line shows a single order condition (red = random, green = recency sorted). Filled symbols denote regression fits which are significantly better than an equally weighted model. (b) Shows the same type of plot for the averaged weighting coefficients across observers. (c) Plots the correlation between the variance in the observers weightings and their discrimination threshold.

Threshold analysis

To examine the effect of an observers' weighting profile on their averaging capability, we calculated observers' discrimination thresholds for discriminating between heads rotated, on average, to the left or to the right. This was done using a method of constant stimuli and psychometric function fitting. Observers' pointer responses were converted to rightwards or leftwards depending on whether they were positive or negative. A cumulative Gaussian function was fit to the proportion of "rightwards" responses using a maximum likelihood method, for each of the mean directions tested. The variance parameter of this fitted Gaussian provides an estimate of the observer's threshold.

In order to determine if unequal weighting of sequential elements leads to poorer averaging performance, a correlation analysis was carried out between each participant's discrimination threshold in the random condition and the variance in their regression weights, where zero variance means all samples are weighted equally. The correlation showed no relationship between these two variables (fig 4c), suggesting that failing to weight all items equally does not limit averaging performance.

Figure 5a shows the averaged thresholds plotted for each participant for the two order conditions. Overall it can be seen that there is no difference in the mean discrimination threshold for the two conditions. Given that there was a large amount of variance in the weighting profiles obtained from the participants, a second analysis was conducted to see if there was a relationship between the size of each participant's recency effect and the difference in threshold between the sorted/random conditions. Each participant's "recency" was calculated as the difference between the mean of the final element, compared to the mean of the other seven in the random condition. This was found to be significantly correlated with the difference in observers' thresholds for the two order conditions (fig 5b). The larger the recency effect, the more the participant benefitted from the change in the sorted order.

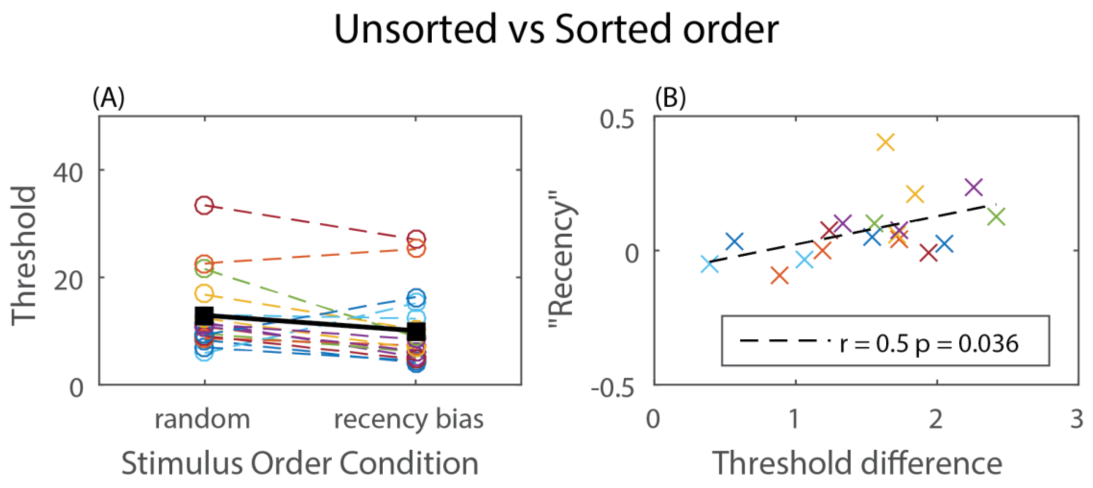

Figure 5:5 Threshold comparisons for sorted and unsorted stimuli.

– Left: Each observer's (different colours) discrimination threshold is plotted for the two order conditions. The mean for each condition is shown in black. Right: A correlation plot showing the relationship between the differences in the threshold for each observer between the two conditions and their "recency" as measured by the difference between the final element and the mean of the other seven. Colours correspond to one observer in both panels.

General Discussion

We used a head rotation averaging task to examine the influence of weighting strategy on averaging performance. The key findings are: (1) we replicate previous findings of recency biases in sequential averaging, though these are smaller than previously observed and do not peak with the final element. (2) Observers' thresholds do not relate to the amount of variance in the weightings of each element, suggesting that unequal weighting of stimuli does not explain inefficient averaging effects previously observed. (3) Observers do not change their strategy when the stimuli are biased to optimise a particular weighting profile, in this case recency. (4) Observers who show a recency bias in a randomly ordered sequence, improve their averaging when the stimuli sequences are sorted to favour this type of bias.

The recency effects we observe here are consistent with previous findings for size averaging (Gorea et al., 2014) and face emotion averaging (Hubert-Wallander & Boynton, 2015). Here we show data from each observer where it can be clearly seen that there are some large inter-observer differences. Although most observers showed a small recency effect, which was a significantly better predictor of their responses than a flat weighting, some observers appeared to equally weight every element whereas others were extremely biased for only the last item in the stimuli. When we take an average across participants we find a small recency effect overall, though this is mainly driven by a sub-set of participants with strong recency biases. It is worth considering in future work looking at weightings in sequential averaging, that presenting mean weightings may dilute potentially important effects within individuals' data.

Different explanations for recency effects have been proposed. One possibility is that visual memory (e.g. (Hay, Smyth, Hitch, & Horton, 2007; Ward, Avons, & Melling, 2005) limits the information that can be used, where only late items are remembered accurately after a sequence is presented. Alternatively, an adaptive gain control model (as proposed by Cheadle et al., 2014), would suggest that observers should update the weighting they give to each item based on how predictable it is. That is, how close it is to the mean of what has been presented to that point. If preceding elements do predict future ones (which they will if all samples come from a fixed distribution) then later items will be up-weighted. It is likely that a combination of these effects play a role in the different weightings given to items in a sequence and it may be that the individual differences we observe are a result of how much these factors affect each individual.

Juni et al. (2012) have previously observed that observers were able to adapt the weightings they gave to different elements in a sequence, based on how precisely each element related to the underlying mean. It is possibly surprising then that we find no difference in averaging strategy when we manipulated our stimuli such that the elements closest to the mean always appeared toward the end of the stimulus. In Juni et al.'s study, observers were given explicit feedback as to what the true mean was on each trial, whereas no feedback was given in the present study. It is likely that through observing which elements were consistent with the true mean over a number of trials, observers could change their weighting to use elements which they knew would give them a reliably correct response. The robustness of observers' strategy in our study suggests that observers cannot implicitly detect differences in the underlying ensemble distribution (or if they do, they do not use them) and that explicit feedback is necessary for them to change how they weight the elements.

Earlier we had hypothesised that sub-sampling, which has been observed in sequential averaging in the past (Gorea et al., 2014), may be the result of un-equal weighting of elements in a sequence and that observers who weighted all items equally should be the most efficient (and therefore accurate) averagers. This was shown not to be the case in the present study; we found no relationship between the amount of variance in the observers' weightings and their discrimination thresholds. Clearly the limit causing sub-sampling is not the weighting profile of the elements and must be linked to the integration of those elements into a summary statistic. What this limit is, and at what point in the averaging process it happens, is still not clear.

Although the variance in observers' weightings did not affect their discrimination thresholds, we have shown that by manipulating the stimuli to target a certain type of bias (recency), we can induce a small but significant improvement in performance. This demonstrates that these biases are having at least some impact on how well observers are averaging, opening up the possibility for further biases to be explored and stimuli designed to best exploit these unconscious biases. It would also be interesting to see whether similar biases that have been observed in spatial averaging (e.g. central biases and robust averaging) could be exploited to improve averaging performance in these tasks.

Chapter 6 Summary and Implications

This thesis examined how cues to social attention, specifically gaze deviation and head rotation, are perceived beyond fixation. Psychophysics experiments and modelling were used to examine social cue processing for faces in the visual periphery as well as groups of faces. The results of these studies show that processing of peripheral gaze deviation is severely limited and heavily influenced by head rotation and that observers struggle to accurately judge the average gaze deviation of a group of faces. The key findings and their implications will be summarised and discussed here.

Peripheral Gaze Perception

The first aim of the thesis was to understand how people perceive gaze direction outside of central fixation, with a particular focus on changes in the "cone of direct gaze". Previous research had presented gaze stimuli at fixation, whereas in day to day life faces will often appear in our periphery, so we must have mechanisms for processing gaze deviation and head rotation under these conditions. The first key finding was that gaze processing is not simply noisier in the periphery compared to at fixation; there are qualitatively differences in the way that gaze is processed outside of fixation.

In the periphery, observers were heavily biased to respond that a face was "looking at them" if the head rotation was also forward facing and biased against saying a face was "looking at them" if the head rotation was averted. A simple explanation for this may have been that observers ignore gaze deviation in the periphery because they cannot process it accurately, and instead just report the head rotation. This basic explanation however does not account for the fact that observers were able to discriminate between leftwards and rightwards gaze deviations in the periphery, and that head rotation was far less influential on these left/right judgements. A more complete explanation is that gaze deviation information is still available in the periphery, and can be used to determine the actual direction that a person is looking in (although with less precision than in the fovea); however, when judging whether or not we are being looked at, head rotation dominates this judgement.

These peripheral gaze processing results have implications for future social cue perception research and highlight further questions yet to be answered. First, our results suggest that the decision the participant has to make when judging the gaze deviation is very important. In this study, participants had to classify gaze as leftwards, direct or rightwards, whereas a number of previous studies (Palanica & Itier, 2015; Yokoyama et al., 2014) have only asked observers to judge between direct or averted gaze. Including the leftwards and rightwards options not only allows the cone of direct gaze model (Calder et al., 2008; Jun et al., 2013) to be fitted; it also has the potential to provide insight into whether observers are using separate information to make judgements about *being looked at* and exactly where another person is looking using a single experimental design.

Second, these results show that the way that social cues (i.e. head rotation and gaze deviation) are combined into a single percept, can vary depending on the conditions under which they are viewed. This implies there must be some high level mechanism which can combine these cues and is sensitive to changes in viewing conditions; further understanding how such a mechanism may operate and how additional social cues like body rotation may be included into judgements is a potential future area of research.

Averaging Groups of Faces

The second aim of this thesis was to investigate how observers judge the average direction of attention from groups of faces; either from gaze deviation or head rotation. Previous research looking at faces had suggested that gaze deviation could be averaged (Sweeny & Whitney, 2014), though only in sets of up to four faces and with no comparison as to how well gaze was averaged relative to head rotation. Studies in this thesis used *equivalent noise analysis*, a method which had previously been used in averaging research to show that, although observers can average visual properties (Dakin, 2001; Dakin, Mareschal, & Bex, 2005), they do not efficiently combine all the elements they are presented with, instead sub-sampling from an array to estimate the average. Estimates of *internal noise* (the uncertainty for judging an individual face in a set) and *effective sample size* (the number of samples observers used to calculate their average) were collected for averaging gaze deviation and head rotation, both when all faces were presented simultaneously and when they were presented in a sequence.

The main finding from the spatial averaging data was that averaging of gaze deviation was much more limited than head rotation, with observers only using between 1 and 2 faces from a set of 16 gaze deviation stimuli in the first experiment (chapter 2). This number did increase to up to 2/3 out of 8 when the size of the faces increased in chapter 3, however the number of faces used for averaging was always less than that for averaging head rotation. Given the results for the peripheral processing of gaze, these results were not necessarily surprising as processing a group of faces requires most of the faces in the set to be processed in the periphery, where perception is known to be limited. Reverse correlation analysis of this data further highlighted this point, as it showed that participants were making their judgements based on faces in the centre of the crowds, an effect that was more pronounced for gaze deviation than head rotation. Taken together, these results suggest that we do not use gaze deviation to rapidly judge the direction of attention of a crowd. It is more likely that the average head rotation is estimated and used to judge what a crowd may be interested in.

In contrast to spatial averaging, comparing averaging performance for gaze and heads over time did not reveal a difference between the two cue types. This lends further strength to the argument that the limiting factor in averaging crowds of gaze stimuli was poor peripheral perception of gaze, as no faces were presented peripherally in the sequential task. When processing groups of faces, head rotation may be used to rapidly extract the average direction of attention and a more precise average is then taken by making a saccade to each face and forming an average of the gaze deviations based on these non-peripheral estimates.

Equivalent Noise Analysis

The equivalent noise analysis that was used in this thesis allowed us to go further than most previous research investigating face averaging had been able to. Where previous research had, for the most part, been focused on showing simply that we *can* average a certain property of faces (e.g. (Haberman et al., 2009; Haberman & Whitney, 2009), the work looks at what limits averaging performance, as well as provides a measure for comparing averaging under different conditions. Future averaging research could benefit from employing these techniques to gain greater insight into averaging mechanisms.

Although EQN does have many benefits for studying averaging, it does not provide us with a full picture of what is occurring when observers are doing an averaging task. The key word in the *effective sample size* parameter is *effective*; EQN can tell us how many samples an observer appears to be using, assuming they are performing the averaging task in an "ideal" way (taking an equally weighted linear average of a sub-set of the samples). Evidence from chapter 4, where regression analysis was used to show that observers are biased to weight later items in a sequence more highly in their average than others, suggests that this is not always the case. Addressing exactly what mechanism is being employed by an observer when they are averaging is a question for future research.

Strategy and Individual Differences in Averaging

Throughout this thesis, individual differences have been observed in participants' ability to average (chapters 2/3) and the strategies they adopt when averaging (chapter 4). These differences were revealed through differences in observers' effective sample size for averaging over space and time, with some observers able to integrate more samples into their average estimates than others. Notably, those who had high ESSs for spatial averaging did not necessarily have high ESSs for sequential averaging, suggesting that these two types of averaging are conducted by separate neural mechanisms. In chapter 4, the weighting that observers gave to each element in a sequence was analysed using a regression model. The results showed an overall recency effect, where observers were biased to weight elements that appear later in a sequence higher than those earlier on. Although this effect was found for the averaged data across participants, examining data for each individual revealed large individual differences in the strategy adopted for combining information over time.

The question of what makes a "good averager" (i.e. somebody who can accurately average arrays with high variance using a large number of samples) has not effectively been answered in this thesis. It seems that being an effective averager transfers between stimulus types, as demonstrated by the correlation between participants effective samples size for averaging head rotation and gaze deviation, though what shared mechanism is limiting this process has not been identified. It was expected that by estimating the area of a spatial array that an observer was using in their average calculation we would reveal differences between good and bad averagers, such as a wider spread of attention for better averagers; unfortunately this was not the case. Similarly, by estimating the weighting applied to each element in a sequence, we hypothesised that better averagers would weight all the items equally, whereas poor averagers would be more subject to biases; again this was not the case. From the data, it could be suggested that the high ESSs for the author (JF) indicate a role of practice and experience in producing better averaging performance. It is possible that practice may have some marginal impact on averaging performance, though the poorer averaging of a co-author for much of this work (IM) suggest this cannot explain the amount of variance in ESSs we observe in this thesis. Clearly voluntary averaging is complex process; further work is needed to understand what neural mechanisms are responsible for the averaging process and what makes some people better averagers than others.

Chapter 7 References

Adams Jr., R. B., & Kleck, R. E. (2005). Effects of Direct and Averted Gaze on the Perception of Facially Communicated Emotion. *Emotion*, *5*(1), 3–11. https://doi.org/10.1037/1528-3542.5.1.3

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, *74*(5), 810–815. https://doi.org/10.3758/s13414-012-0293-0

Allard, R., & Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *Journal of Vision*, *12*(11), 6–6. https://doi.org/10.1167/12.11.6

Allen, H. A., Hess, R. F., Mansouri, B., & Dakin, S. C. (2003). Integration of first- and second-order orientation. *Journal of the Optical Society of America A*, *20*(6), 974. https://doi.org/10.1364/JOSAA.20.000974

Alvarez. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Anstis, S. M., Mayhew, J. W., & Morley, T. (1969). The Perception of Where a Face or Television 'Portrait' Is Looking. *The American Journal of Psychology*, *82*(4), 474. https://doi.org/10.2307/1420441

Ariely, D. (2001). Seeing Sets: Representation by Statistical Properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, *77*(4), 1116–1131. https://doi.org/10.3758/s13414-015-0870-0

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of Physiology*, *136*(3), 469–488.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional Facial Expression Detection in the Peripheral Visual Field. *PLoS ONE*, *6*(6), e21584. https://doi.org/10.1371/journal.pone.0021584

Bayliss, A. P., & Tipper, S. P. (2005). Gaze and arrow cueing of attention reveals individual differences along the autism spectrum as a function of target context. *British Journal of Psychology*, *96*(1), 95–114. https://doi.org/10.1348/000712604X15626

Bock, S. W., Dicke, P., & Thier, P. (2008). How precise is gaze following in humans? *Vision Research*, *48*(7), 946–957. https://doi.org/10.1016/j.visres.2008.01.011

Bossi, M., & Dakin, S. C. (2014). Voluntary averaging of two-dimensional spatial position. *Perception*, *43*, 43.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brand, J., Oriet, C., & Tottenham, L. S. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(1), 63–69. https://doi.org/10.1037/a0026950

Calder, A. J., Beaver, J. D., Winston, J. S., Dolan, R. J., Jenkins, R., Eger, E., & Henson, R. N. A. (2007). Separate Coding of Different Gaze Directions in the Superior Temporal Sulcus and Inferior Parietal Lobule. *Current Biology*, *17*(1), 20–25. https://doi.org/10.1016/j.cub.2006.10.052

Calder, A. J., Jenkins, R., Cassel, A., & Clifford, C. W. G. (2008). Visual representation of eye gaze is coded by a nonopponent multichannel system. *Journal of Experimental Psychology: General*, *137*(2), 244–261. https://doi.org/10.1037/0096-3445.137.2.244

Calvo, M. G., Fernández-Martín, A., & Nummenmaa, L. (2014). Facial expression recognition in peripheral versus central vision: role of the eyes and the mouth. *Psychological Research*, *78*(2), 180–195. https://doi.org/10.1007/s00426-013-0492-x

Calvo, M. G., & Lang, P. J. (2005). Parafoveal Semantic Processing of Emotional Visual Scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 502–519. https://doi.org/10.1037/0096-1523.31.3.502

Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*(11), 1123–1142. https://doi.org/10.1016/0028-3932(90)90050-X

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S., & Summerfield, C. (2014). Adaptive Gain Control during Human Perceptual Choice. *Neuron*, *81*(6), 1429–1441. https://doi.org/10.1016/j.neuron.2014.01.020

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13. https://doi.org/10.3758/BF03195009

Chong, S. C., & Treisman, A. (2005b). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

Chung, S. T. L., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, *38*(19), 2949–2962. https://doi.org/10.1016/S0042-6989(98)00072-8

Cline, M. G. (1967). The Perception of Where a Person Is Looking. *The American Journal of Psychology*, *80*(1), 41. https://doi.org/10.2307/1420539

Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *The Journal of Comparative Neurology*, *300*(1), 5–25. https://doi.org/10.1002/cne.903000103

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*(5), 1016–1026. https://doi.org/10.1364/JOSAA.18.001016

Dakin, S. C. (2015). Seeing statistical regularities. In *Oxford Handbook of Perceptual Regularities*. Oxford, UK: Oxford University Press.

Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness 'filling in'. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1531), 2341–2348. https://doi.org/10.1098/rspb.2003.2528

Dakin, S. C., Cass, J., Greenwood, J. A., & Bex, P. J. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *Journal of Vision*, *10*(10), 14. https://doi.org/10.1167/10.10.14

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, *45*(24), 3027–3049. https://doi.org/10.1016/j.visres.2005.07.037

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/S0042-6989(97)00133-8

Dakin, S. C., & Watt, R. J. (2009). Biological 'bar codes' in human faces. *Journal of Vision*, *9*(4), 2. https://doi.org/10.1167/9.4.2

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, *20*(4), 569–572. https://doi.org/10.1016/S0163-6383(97)90048-5

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition*, *6*(5), 509–540. https://doi.org/10.1080/135062899394920

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The Cost of Accumulating Evidence in Perceptual Decision Making. *The Journal of Neuroscience*, *32*(11), 3612–3628. https://doi.org/10.1523/JNEUROSCI.4010-11.2012

Duncan, R. O., & Boynton, G. M. (2003). Cortical Magnification within Human Primary Visual Cortex Correlates with Acuity Thresholds. *Neuron*, *38*(4), 659–671. https://doi.org/10.1016/S0896-6273(03)00265-4

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Ewbank, M. P., Jennings, C., & Calder, A. J. (2009). Why are you angry with me? Facial expressions of threat influence perception of gaze direction. *Journal of Vision*, *9*(12), 16. https://doi.org/10.1167/9.12.16

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Farroni, T., Massaccesi, S., Menon, E., & Johnson, M. H. (2007). Direct gaze modulates face recognition in young infants. *Cognition*, *102*(3), 396–404. https://doi.org/10.1016/j.cognition.2006.01.007

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*. https://doi.org/10.1038/srep32210

Florey, J., Dakin, S. C., Clifford, C. W. G., & Mareschal, I. (2015). Peripheral processing of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1084–1094. https://doi.org/10.1037/xhp0000068

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6–6. https://doi.org/10.1167/8.2.6

Franck, N., Daprati, E., Michel, F., Saoud, M., Daléry, J., Marie-Cardine, M., & Georgieff, N. (1998). Gaze discrimination is unimpaired in schizophrenia. *Psychiatry Research*, *81*(1), 67–75. https://doi.org/10.1016/S0165-1781(98)00082-1

Franck, N., Montoute, T., Labruyère, N., Tiberghien, G., Marie-Cardine, M., Daléry, J., … Georgieff, N. (2002). Gaze direction determination in schizophrenia. *Schizophrenia Research*, *56*(3), 225–234. https://doi.org/10.1016/S0920-9964(01)00263-8

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, *5*(3), 490–495. https://doi.org/10.3758/BF03208827

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences*, *109*(19), 7245–7250. https://doi.org/10.1073/pnas.1116141109

Gamer, M., & Hecht, H. (2007). Are you looking at me? Measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 705–715. https://doi.org/10.1037/0096-1523.33.3.705

Gibson, J. J., & Pick, A. D. (1963). Perception of Another Person's Looking Behavior. *The American Journal of Psychology*, *76*(3), 386. https://doi.org/10.2307/1419779

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, *14*(9), 22–22. https://doi.org/10.1167/14.9.22

Grice, S. J., Halit, H., Farroni, T., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2005). Neural Correlates of Eye-Gaze Detection in Young Children with Autism. *Cortex*, *41*(3), 342–353. https://doi.org/10.1016/S0010-9452(08)70271-5

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1–1. https://doi.org/10.1167/9.11.1

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: a SIMPLE explanation? *Memory & Cognition*, *35*(1), 176–190.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106. https://doi.org/10.3758/BF03195300

Hock, H. S., & Schöner, G. (2010). Measuring Perceptual Hysteresis with the Modified Method of Limits: Dynamics at the Threshold. *Seeing and Perceiving*, *23*(2), 173–195. https://doi.org/10.1163/187847510X503597

Hood, B. M., Macrae, C. N., Cole-Davies, V., & Dias, M. (2003). Eye remember you: the effects of gaze direction on face recognition in children and adults. *Developmental Science*, *6*(1), 67–71. https://doi.org/10.1111/1467-7687.00256

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's Eyes Trigger Shifts of Visual Attention in Human Infants. *Psychological Science*, *9*(2), 131–134. https://doi.org/10.1111/1467-9280.00024

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision*, *15*(4). https://doi.org/10.1167/15.4.5

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Jenkins, J., & Langton, S. R. H. (2003). Configural processing in the perception of eye-gaze direction. *Perception*, *32*(10), 1181–1188.

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I Thought You Were Looking at Me Direction-Specific Aftereffects in Gaze Perception. *Psychological Science*, *17*(6), 506–513. https://doi.org/10.1111/j.1467-9280.2006.01736.x

Jun, Y. Y., Mareschal, I., Clifford, C. W. G., & Dadds, M. R. (2013). Cone of direct gaze as a marker of social anxiety in males. *Psychiatry Research*, *210*(1), 193–198. https://doi.org/10.1016/j.psychres.2013.05.020

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2012). Effective integration of serially presented stochastic cues. *Journal of Vision*, *12*(8), 12–12. https://doi.org/10.1167/12.8.12

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A*, *4*(12), 2395. https://doi.org/10.1364/JOSAA.4.002395

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. https://doi.org/10.1037/0033-2909.100.1.78

Kliemann, D., Dziobek, I., Hatri, A., Steimke, R., & Heekeren, H. R. (2010). Atypical Reflexive Gaze Patterns on Emotional Faces in Autism Spectrum Disorders. *The Journal of Neuroscience*, *30*(37), 12281–12287. https://doi.org/10.1523/JNEUROSCI.0688-10.2010

Langton, S. R. H., Honeyman, H., & Tessler, E. (2004). The influence of head contour and nose angle on the perception of eye-gaze direction. *Perception & Psychophysics*, *66*(5), 752–771. https://doi.org/10.3758/BF03194970

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Lawson, R. P., Clifford, C. W. G., & Calder, A. J. (2011). A real head turner: Horizontal and vertical head directions are multichannel coded. *Journal of Vision*, *11*(9), 17–17. https://doi.org/10.1167/11.9.17

Leekam, S. R., Hunnisett, E., & Moore, C. (1998). Targets and Cues: Gaze-following in Children with Autism. *Journal of Child Psychology and Psychiatry*, *39*(7), 951–962. https://doi.org/10.1111/1469-7610.00398

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial Dependence in the Perception of Faces. *Current Biology*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*(9), 1443 – 1457. https://doi.org/10.1068/p5896

Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., & Mason, M. F. (2002). Are You Looking at Me? Eye Gaze and Person Perception. *Psychological Science*, *13*(5), 460–464. https://doi.org/10.1111/1467-9280.00481

Mäkelä, P., Näsänen, R., Rovamo, J., & Melmoth, D. (2001). Identification of facial images in peripheral vision. *Vision Research*, *41*(5), 599–610. https://doi.org/10.1016/S0042-6989(00)00259-5

Manning, C., Dakin, S. C., Tibber, M. S., & Pellicano, E. (2014). Averaging, not internal noise, limits the development of coherent motion processing. *Developmental Cognitive Neuroscience*, *10*, 44–56. https://doi.org/10.1016/j.dcn.2014.07.004

Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2013). Humans Have an Expectation That Gaze Is Directed Toward Them. *Current Biology*, *23*(8), 717–721. https://doi.org/10.1016/j.cub.2013.03.030

Mareschal, I., Calder, A. J., Dadds, M. R., & Clifford, C. W. G. (2013). Gaze categorization under uncertainty: Psychophysics and modeling. *Journal of Vision*, *13*(5), 18. https://doi.org/10.1167/13.5.18

Mareschal, I., & Clifford, C. W. G. (2012). Dynamics of unconscious contextual effects in orientation processing. *Proceedings of the National Academy of Sciences*, *109*(19), 7553–7558. https://doi.org/10.1073/pnas.1200952109

Mareschal, I., Dakin, S. C., & Bex, P. J. (2006). Dynamic properties of orientation discrimination assessed by using classification images. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 5131–5136. https://doi.org/10.1073/pnas.0507259103

Mareschal, I., Morgan, M. J., & Solomon, J. A. (2010). Cortical distance determines whether flankers cause crowding or the tilt illusion. *Journal of Vision*, *10*(8), 13. https://doi.org/10.1167/10.8.13

Mareschal, I., Otsuka, Y., & Clifford, C. W. G. (2014). A generalized tendency toward direct gaze with uncertainty. *Journal of Vision*, *14*(12), 27–27. https://doi.org/10.1167/14.12.27

Mareschal, I., Otsuka, Y., & Clifford, C. W. G. (2014). A generalized tendency toward direct gaze with uncertainty. *Journal of Vision*, *14*(12), 27. https://doi.org/10.1167/14.12.27

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, *13*(2), 79–82. https://doi.org/10.1037/h0028070

Morrone, M. C., Burr, D. C., Pietro, S. D., & Stefanelli, M.-A. (1999). Cardinal directions for visual optic flow. *Current Biology*, *9*(14), 763–766. https://doi.org/10.1016/S0960-9822(99)80338-8

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. https://doi.org/10.3758/PP.70.5.772

Näsänen, R., & O'Leary, C. (1998). Recognition of band-pass filtered hand-written numerals in foveal and peripheral vision. *Vision Research*, *38*(23), 3691–3701. https://doi.org/10.1016/S0042-6989(98)00062-5

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*(6754), 695–698. https://doi.org/10.1038/44409

Neumann, D., Spezio, M. L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience*, *1*(3), 194–202. https://doi.org/10.1093/scan/nsl030

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8–16. https://doi.org/10.1016/j.visres.2012.12.004

Otsuka, Y., Mareschal, I., Calder, A. J., & Clifford, C. W. G. (2014). Dual-Route Model of the Effect of Head Orientation on Perceived Gaze Direction. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/a0036151

Palanica, A., & Itier, R. J. (2015). Effects of peripheral eccentricity and head orientation on gaze discrimination. *Visual Cognition*, *22*(9–10), 1216–1232. https://doi.org/10.1080/13506285.2014.990545

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. https://doi.org/10.1038/89532

Pelli, D., & Blakemore, C. (1990). The quantum efficiency of vision. *Vision: Coding and Efficiency*, 3–24.

Perrett, D. I., Smith, P. a. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *223*(1232), 293–317. https://doi.org/10.1098/rspb.1985.0003

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans Use Summary Statistics to Perceive Auditory Sequences. *Psychological Science*, *24*(8), 1389–1397. https://doi.org/10.1177/0956797612473759

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Ricciardelli, P., Baylis, G., & Driver, J. (2000). The positive and negative of human expertise in gaze perception. *Cognition*, *77*(1), B1–B14. https://doi.org/10.1016/S0010-0277(00)00092-5

Ricciardelli, P., & Driver, J. (2008). Effects of head orientation on gaze perception: How positive congruency effects can be reversed. *The Quarterly Journal of Experimental Psychology*, *61*(3), 491–504. https://doi.org/10.1080/17470210701255457

Rimmele, U., & Lobmaier, J. S. (2012). Stress increases the feeling of being looked at. *Psychoneuroendocrinology*, *37*(2), 292–298. https://doi.org/10.1016/j.psyneuen.2011.06.013

Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*(2), 147–166. https://doi.org/10.1016/j.cogsci.2003.11.003

Rosse, R. B., Kendrick, K., Wyatt, R. J., Isaac, A., & Deutsch, S. I. (1994). Gaze discrimination in patients with schizophrenia: preliminary report. *American Journal of Psychiatry*, *151*(6), 919–921.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, *37*(3), 495–510. https://doi.org/10.1007/BF00236819

Rovamo, J., Virsu, V., Laurinen, P., & Hyvärinen, L. (1982). Resolution of gratings oriented along and across meridians in peripheral vision. *Investigative Ophthalmology & Visual Science*, *23*(5), 666–670.

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition*, *12*(1), 127–144. https://doi.org/10.1080/13506280444000157

Senju, A., & Johnson, M. H. (2009a). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1204–1214. https://doi.org/10.1016/j.neubiorev.2009.06.001

Senju, A., & Johnson, M. H. (2009b). The eye contact effect: mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. https://doi.org/10.1016/j.tics.2008.11.009

Senju, A., Tojo, Y., Yaguchi, K., & Hasegawa, T. (2005). Deviant gaze processing in children with autism: an ERP study. *Neuropsychologia*, *43*(9), 1297–1306. https://doi.org/10.1016/j.neuropsychologia.2004.12.002

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14), 19–19. https://doi.org/10.1167/10.14.19

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14), 19. https://doi.org/10.1167/10.14.19

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 13. https://doi.org/10.1167/11.12.13

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 13–13. https://doi.org/10.1167/11.12.13

Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention Ensemble Perception of a Crowd’s Gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, *45*(5), 643–659. https://doi.org/10.1016/j.visres.2004.09.017

Todorović, D. (2009). The effect of face eccentricity on the perception of gaze direction. *Perception*, *38*(1), 109–132. https://doi.org/10.1068/p5930

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 4–4. https://doi.org/10.1167/9.7.4

Vecera, S. P., & Rizzo, M. (2006). Eye gaze does not produce reflexive shifts of attention: Evidence from frontal-lobe damage. *Neuropsychologia*, *44*(1), 150–159. https://doi.org/10.1016/j.neuropsychologia.2005.04.010

Vida, M. D., & Maurer, D. (2012). The development of fine-grained sensitivity to eye contact after 6 years of age. *Journal of Experimental Child Psychology*, *112*(2), 243–256. https://doi.org/10.1016/j.jecp.2012.02.002

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, *37*(3), 475–494. https://doi.org/10.1007/BF00236818

Vuilleumier, P., George, N., Lister, V., Armony, J., & Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Visual Cognition*, *12*(1), 85–101. https://doi.org/10.1080/13506280444000120

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: functional equivalence across modalities. *Memory (Hove, England)*, *13*(3–4), 308–317.

Webster, M. A. (1996). Human colour perception and its adaptation. *Network: Computation in Neural Systems*, *7*(4), 587–634. https://doi.org/10.1088/0954-898X_7_4_002

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety?—An eye-tracking and psychophysiology study. *Journal of Anxiety Disorders*, *23*(1), 93–103. https://doi.org/10.1016/j.janxdis.2008.04.004

Wilson, H. R., Wilkinson, F., Lin, L.-M., & Castillo, M. (2000). Perception of head orientation. *Vision Research*, *40*(5), 459–472. https://doi.org/10.1016/S0042-6989(99)00195-9

Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, *15*(4), 11. https://doi.org/10.1167/15.4.11

Wollaston, W. H. (1824). On the Apparent Direction of Eyes in a Portrait. *Philosophical Transactions of the Royal Society of London*, *114*, 247–256. https://doi.org/10.1098/rstl.1824.0016

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In *Eye Movements and Vision* (pp. 171–211). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4899-5379-7_8

Yokoyama, T., Sakai, H., Noguchi, Y., & Kita, S. (2014). Perception of Direct Gaze Does Not Require Focus of Attention. *Scientific Reports*, *4*. https://doi.org/10.1038/srep03858